PNAS RESEARCH ARTICLE | EVOLUTION

A method to predict the response to directional selection using a Kalman filter

Lisandro Milocco[a,1] and Isaac Salazar-Ciudad[a,b,c,1]

Edited by Nils Stenseth, Universitetet i Oslo, Oslo, Norway; received September 29, 2021; accepted May 2, 2022

Predicting evolution remains challenging. The field of quantitative genetics provides predictions for the response to directional selection through the breeder's equation, but these predictions can have errors. The sources of these errors include omission of traits under selection, inaccurate estimates of genetic variance, and nonlinearities in the relationship between genetic and phenotypic variation. Previous research showed that the expected value of these prediction errors is often not zero, so predictions are systematically biased. Here, we propose that this bias, rather than being a nuisance, can be used to improve the predictions. We use this to develop a method to predict evolution, which is built on three key innovations. First, the method predicts change as the breeder's equation plus a bias term. Second, the method combines information from the breeder's equation and from the record of past changes in the mean to predict change using a Kalman filter. Third, the parameters of the filter are fitted in each generation using a learning algorithm on the record of past changes. We compare the method to the breeder's equation in two artificial selection experiments, one using the wing of the fruit fly and another using simulations that include a complex mapping of genotypes to phenotypes. The proposed method outperforms the breeder's equation, particularly when traits under selection are omitted from the analysis, when data are noisy, and when additive genetic variance is estimated inaccurately or not estimated at all. The proposed method is easy to apply, requiring only the trait means over past generations.

quantitative genetics | evolutionary prediction | Kalman filter | breeder's equation | G matrix

Significance

Evolution is a historical process with contingency, where outcomes are sensitive to past events. Due to this, predicting evolution remains challenging. In this paper, we propose a method to predict the response to selection that incorporates history. The method uses tools from quantitative genetics and combines them with information from past evolution that is extracted through a combination of signal processing and learning techniques. This information is underexploited by existing methods to predict evolution but is of great value since it reflects singularities of the evolutionary system. We show that this combination of information coming from the time series and quantitative genetics methods outperforms classical methods in predicting the response to selection.

Evolutionary prediction is an important and active field within evolutionary biology (1–4). Aside from its theoretical value, predicting evolution has important applications, such as developing strategies for the persistence of populations amid rapid environmental change (5) and designing interventions to control the spread of a disease (6).

Quantitative genetics is a widely used approach to study and predict short-term evolution of continuous traits (7, 8). The backbone of this theory is the breeder's equation (9, 10). In its multivariate form, it provides predictions of the change in the mean of a set of traits, from one generation to the next, in response to directional selection:

$$\boldsymbol{\Delta}\bar{\boldsymbol{z}}_i = G_i P_i^{-1} \boldsymbol{s}_i, \qquad [1]$$

where $\boldsymbol{\Delta}\bar{\boldsymbol{z}}_i = \bar{\boldsymbol{z}}_{i+1} - \bar{\boldsymbol{z}}_i$ is the vector of change in trait means from generation i to $i+1$; G_i and P_i are additive genetic and phenotypic variance–covariance matrices between traits in generation i, respectively; and $\boldsymbol{s}_i$ is the vector of selection differentials in generation i.

A major appeal of the equation is that its elements can be estimated without detailed knowledge of the genetic architecture and development underlying the focal traits. Indeed, estimates of G_i and P_i can be obtained using only phenotypic data and known genetic relatedness among individuals in a population (11, 12), while estimates of $\boldsymbol{s}_i$ need knowledge of individual fitness (8, 13). The simplicity of the equation, however, is achieved at the cost of some assumptions.

The breeder's equation assumes an infinitesimal model of genetic effects (i.e., a large number of loci, each of small effect) or at least a linear parent–offspring regression (and a few additional assumptions) (8, 14). It further assumes linkage equilibrium, Hardy–Weinberg proportions, random associations of environmental effects, and that the traits being analyzed are uncorrelated with other unmeasured traits under selection (2, 8, 13). Moreover, the equation is local, meaning that the accuracy of the predictions can only be ensured for a single generation (8).

When applied to real systems, the assumptions of the breeder's equation are violated to some extent. A common violation is the so-called missing character problem, where the particular traits chosen for study do not account for all selection (2, 15). Another violation arises when the equation is used to predict the response for several generations under

Author affiliations: [a]Institute of Biotechnology, Helsinki Institute of Life Science, University of Helsinki, 00014 Helsinki, Finland; [b]Centre de Recerca Matemàtica, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain; and [c]Genomics, Bioinformatics and Evolution, Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Barcelona, Spain

Author contributions: L.M. conceptualized and developed the method, with input from I.S.-C.; L.M. performed the research and wrote the original draft; and L.M. and I.S.-C. designed the experiments and edited the original draft.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: milocco.lisandro@gmail.com or isaac.salazar@uab.cat.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2117916119/-/DCSupplemental.

Published July 6, 2022.

the assumption that the statistical parameters remain constant over these generations. However, the constancy of the G matrix is a debated issue (16, 17) since it can be affected by mutation, drift, and selection itself (18), and work on nonlinear genotype–phenotype maps (19) and gene–environment interactions (20, 21) shows that the G matrix can change rapidly even in a few generations. This issue is aggravated by the fact that the breeder's equation assumes that the real statistical parameters G_i, P_i, and s_i are known without error; in reality, these are inaccessible parameters that must be replaced by estimates. While this is not strictly a violation of the assumptions, this introduces uncertainty and possibly biases to the predictions, particularly when G_i is estimated and used in different environments (22) or when relevant effects are not controlled for during the estimation of G_i [e.g., maternal effects (7, 8, 15)].

Indeed, when applied to real systems, violations of the assumptions of the breeder's equation can lead to prediction errors (2, 8, 14, 15, 23–25). A notable example is the problem of stasis (2, 26), where no response to selection is observed in a population that both has ample additive genetic variance and is under strong directional selection [e.g., body size in several species, including Soay sheep and red deer (26)]. Prediction errors have also been reported in artificial selection experiments when the parent–offspring regression is nonlinear (23, 27) and when selection is applied in the direction opposite to the sign of the genetic correlation between the two traits under selection (7).

An important feature of the prediction errors when using the breeder's equation is that their mean over time can be nonzero (14, 25), indicating the presence of a systematic bias. For example, if a trait under selection is missing from the analysis, the prediction using the breeder's equation can be biased because there is an indirect effect of selection that is systematically omitted in the prediction (26). Moreover, if the G matrix has changed from its original estimate, predictions will also be biased because the G used for predictions is incorrect. In this way, the total prediction error (i.e., the difference between the prediction of the breeder's equation and the true response to selection) is composed of two parts. One part of the error is stochastic due to drift and measurement noise. The other part is deterministic, a systematic bias. Due to this systematic bias, the error at a given generation i is informative of the error at generation $i + 1$. This indicates that there is potential to improve predictions by incorporating this bias if one could retain the information of past generations as a "memory."

Here, we propose a method to predict the response to directional selection that yields better predictions when some of the assumptions of the breeder's equation do not hold. The method, which we refer to as the KF method, uses the record of the means of the traits in past generations to improve predictions. There are three key innovations in the method. First, it uses a model for the change in the mean of the traits that is the breeder's equation plus a bias term, which is the term with memory. Second, the method predicts the change in the traits and the bias in each generation using a Kalman filter (28). The filter integrates the information of the breeder's equation and the record of past means of the traits, and it efficiently deals with the stochastic component of the prediction error. Third, the method incorporates a learning scheme to learn the parameters of the filter required to provide predictions in each generation.

The Kalman filter is a hallmark of signal processing theory (28, 29) and has a wide variety of technological applications from navigation of aircrafts to econometrics. The filter is a general algorithm that allows the estimation of the value of a set of variables of interest using a model of how the variables are expected to change and a series of measurements observed over time. Here, we adapt it to be used in the prediction of the response to selection.

In *Results*, we first develop the KF method in three parts. *Part I: The Breeder's Equation Plus a Bias Term* is the introduction of the extended equation that consists of the breeder's equation plus a bias term. *Part II: The Kalman Filter* is the development of the Kalman filter for this application. *Part III: Learning the Parameters of the Kalman Filter* is the explanation of the learning algorithm to learn the parameters of the filter at each generation. Later in *Results*, the KF method is used to predict the response to selection in two artificial selection experiments. The datasets are used to explore common situations where the assumptions of the breeder's equation are violated to some extent, as explained above.

Results

Part I: The Breeder's Equation Plus a Bias Term. We want to predict the change in the mean of a set of traits between generations, $\boldsymbol{\Delta}\bar{\boldsymbol{z}}_i$. We propose the following equation consisting of the breeder's equation plus a bias term $\boldsymbol{b}_i$, a vector of length equal to the number of traits:

$$\boldsymbol{\Delta}\bar{\boldsymbol{z}}_i = G_i P_i^{-1} \boldsymbol{s}_i + \boldsymbol{b}_i. \quad \textbf{[2]}$$

The bias term can be understood as the part of the response to selection that is not captured by the breeder's equation and that arises from violations of assumptions. As such, the systematic bias is structured, and we expect the bias of generation $i - 1$ to be similar to the bias at generation i (14, 25).

Here, we propose to estimate the bias term by using measurements of the system up to generation i. In principle, one could estimate the $\boldsymbol{b}_{i-1}$ as the difference between the realized change in the mean, $\boldsymbol{\Delta}\bar{\boldsymbol{z}}_{i-1}$, and the prediction from the breeder's equation, $G_{i-1} P_{i-1}^{-1} \boldsymbol{s}_{i-1}$. Assuming that the bias changes slowly, one could further assume that $\boldsymbol{b}_i \approx \boldsymbol{b}_{i-1}$ and obtain an estimation for the bias at generation i. The problem with this approach is that both the breeder's prediction and the change in the mean for the trait are measured with noise, which typically is very large. This random component of the prediction error is due to several factors, including drift and sampling, and can be so large as to obscure the bias estimated using only the past generation (30). Furthermore, the stochastic component of the error is independent in each generation, so it contains no useful information that can be exploited to improve predictions. This component of the error, therefore, has to be separated from the deterministic part (i.e., the bias), which can be used to improve predictions.

To minimize the stochastic component of the prediction error, we propose here to use a Kalman filter to estimate $\boldsymbol{\Delta}\bar{\boldsymbol{z}}_i$ and $\boldsymbol{b}_i$ in each generation. The filter is explained in the next section. Most importantly, the filter works by efficiently separating the stochastic component of the error from the bias. To simplify the bookkeeping and notation, we will develop the equations of the filter for each trait separately. We then rewrite Eq. **2** for each trait as

$$\Delta_i = \Delta_i^B + b_i, \quad \textbf{[3]}$$

where Δ_i is the change in the mean of a given trait in generation i, Δ_i^B is the prediction using the breeder's equation, and b_i is the bias. In the next section, we show how we use a Kalman filter to estimate the state variables Δ and b in each generation.

Part II: The Kalman Filter. The Kalman filter is a general algorithm to estimate the value of a set of state variables from noisy measurements (29). To achieve this, the filter integrates two sources of information. First, it uses a model of how we expect the state vector to change from one generation to the next. This makes the algorithm recursive since the estimate of the state vector at time $i-1$ is used to make an estimate of the state vector at i. This estimate is combined with a second source of information to make the estimate of the state vector at time i. This second source of information is a set of measurements from the system taken at time i that are related to the state vector. The filter combines these two sources of information by a weighted average. How the average is obtained is the central part of the filter, and it is achieved by calculating a weight matrix that minimizes the error in the estimates (29). Note that both sources of information described above have associated noise summarized by the covariance matrices R_i and Q_i (explained below). These matrices are the parameters of the filter that have to be provided by the user (*Part III: Learning the Parameters of the Kalman Filter*).

For this particular application of the Kalman filter, the state vector at time i is composed of Δ_{i-1} and b_i. Note that with the above definitions, estimating Δ_i gives us a prediction for $\bar{z}_{i+1}$ since $\bar{z}_{i+1} = \bar{z}_i + \Delta_i$. In developing the algorithm below, we use the symbol $\hat{}$ to refer to estimates of the variables (e.g., $\hat{\Delta}_i$ is the estimate of the state variable Δ_i). We make the assumption that the response to directional selection does not show abrupt changes from one generation to the next (8). Additionally, in this application we will assume that the bias changes slowly in time. In this way, we define $\Delta_{i-1} = \Delta_{i-2} + \eta_i$ and $b_i = b_{i-1} + \eta_i^b$, where $\boldsymbol{\eta}_i = (\eta_i, \eta_i^b)$ is a vector of small changes that we assume to be normally distributed with mean zero and covariance matrix Q_i. Note that the assumption here is not that the state variables are constant in time but rather, that they do not show abrupt changes from one generation to the next.

There are two measurements at time i that we can use to improve our estimates. We use the symbol $\sim$ to indicate that the variable has been measured with noise. The measurements are $\tilde{\Delta}_i^B = \Delta_i^B + v_i^B$ and $\tilde{\Delta}_{i-1} = \Delta_{i-1} + v_i$, where we assume that $\boldsymbol{v}_i = (v_i^B, v_i)$ is a vector of Gaussian measurement error with mean zero and covariance matrix R_i.

The Kalman filter combines the estimates of the state variables in $i-1$ (i.e., $\hat{\Delta}_{i-2}$ and $\hat{b}_{i-1}$) and the new measurements (i.e., $\tilde{\Delta}_i^B$ and $\tilde{\Delta}_{i-1}$) to provide the best possible estimates of the state variables in generation i (i.e., $\hat{\Delta}_{i-1}$ and $\hat{b}_i$). Given the relationships described above, this is done using the following formula:

$$\begin{pmatrix} \hat{\Delta}_{i-1} \\ \hat{b}_i \end{pmatrix} = \begin{pmatrix} \hat{\Delta}_{i-2} \\ \hat{b}_{i-1} \end{pmatrix} + K_i \left(\begin{pmatrix} \tilde{\Delta}_i^B \\ \tilde{\Delta}_{i-1} \end{pmatrix} - \begin{pmatrix} \hat{\Delta}_{i-2} - \hat{b}_{i-1} \\ \hat{\Delta}_{i-2} \end{pmatrix} \right). \qquad \textbf{[4]}$$

The first term of the right-hand side is the state vector estimates in step $i-1$. The second term is the correction, which is the product of the matrix K_i and the error. The error is formed by the difference between the measurements $\tilde{\Delta}_i^B$ and $\tilde{\Delta}_{i-1}$ and their expected values using the estimates at step $i-1$ (*Appendix A*). The estimate of the bias is finally used to predict the change at generation i using Eq. **3**, as $\hat{\Delta}_i = \tilde{\Delta}_i^B + \hat{b}_i$.

K_i is a 2×2 matrix called the Kalman gain, which assigns weights to the correction. The calculation of K_i is the key of the filter, and it is done for each i. K_i is a trade-off between the confidence we have on the estimate of the state vector at $i-1$ and the confidence we have on our measurements at generation i, and it is calculated to minimize the error covariance matrix of the estimates (28, 29). If the measurements are to be trusted, then the gain will give more weight to the second term of Eq. **4**. If the estimates at $i-1$ are to be trusted, then the gain will assign more weight to the first term of the equation. The "trust" is quantified by the associated error covariance matrices. This together with the calculation of the gain K_i is explained in *Appendix A*.

As mentioned above, the algorithm is recursive; the estimates obtained in generation $i-1$ using Eq. **4** are the starting point for the prediction in generation i. We then require initial estimates at time $i = 1$ to begin the recursion. For our state variables, $\hat{b}_1 = 0$, and $\hat{\Delta}_0$ is the prediction using the breeder's equation.

Part III: Learning the Parameters of the Kalman Filter. The matrices Q_i and R_i have to be provided by the user to implement the filter explained in *Part II: The Kalman Filter*. Q_i is the covariance of the vector $\boldsymbol{\eta}_i$, and R_i is the covariance of vector $\boldsymbol{v}_i$, which describes measurement noise. These matrices are hard to calculate analytically. For example, the variance in the measurement noise for $\tilde{\Delta}_{i-1}$ is affected by drift, selection, measurement, and sampling (8). An added difficulty is that R_i and Q_i can change in time.

Instead of calculating the matrices R_i and Q_i analytically, we learn them using the time series of the trait means. Several methods exist to identify these matrices using data from the time series (31–33). Here, we propose a simple method to learn the matrices at generation i using a window of the last L recorded changes in the mean $\{\tilde{\Delta}_{i-L}, \ldots, \tilde{\Delta}_{i-1}\}$ to learn the values of R_i and Q_i (similar to the method presented in ref. 33). This is done by running the filter inside the window with several combinations of R_i and Q_i. We then calculate the prediction error of the method in the window for each combination of R_i and Q_i. The error is calculated as the difference between the prediction $\hat{\Delta}_i$ and the true Δ_i. If the trait mean in each generation is measured with error, the true Δ_i can be estimated by first making a linear regression of the means against generations in the window and then, calculating the per-generation change (11). The combination of R_i and Q_i that results in the smallest error is then used to make the actual prediction of interest at time i. Note that this process is done in every generation i for each time series separately. In this way, the method learns the best R_i and Q_i possible for the specific system at time i.

To learn the matrices, we assume that R_i and Q_i are roughly constant inside the window. This sets a limit to how large the window can be since if the window is too large, then the matrices may change substantially inside the window. Then, the size of the window should be kept relatively small, making it hard to learn all the elements of the 2×2 matrices R_i and Q_i (i.e., more elements to learn require a larger dataset). To reduce the number of elements to learn, we make additional simplifications about R_i and Q_i, as detailed in *Appendix B*. These simplifications allow us to reduce the identification to a single parameter to be learned from the data in the moving time window. We call this parameter ρ_i, and it summarizes the relationship between the Q_i and R_i matrices (*Appendix B*). The value of ρ_i will adjust in each generation to reflect the amount of stochastic noise in the data inside the time window, which is of key importance to determine the value of K as explained above. For example, if measurement noise is large, then ρ_i adjusts accordingly by becoming small so that the gain K_i assigns little weight to the second term of Eq. **4**. For the analyses in this paper, we use a window of size L for $i > L$ and size $L = i$ for $2 < i \leqslant L$ (i.e., we use the available generations in the record).

Apart from using the window to learn the parameter ρ_i of the Kalman filter, we also use it to approximate the uncertainty in the predictions using the KF method. To do this, we calculate the SD of the residuals of the predictions against the true change inside the window. We use this as the uncertainty for the predictions of the method.

Testing the Method. We compare the performance of different methods in predicting the response to directional selection using two artificial selection experiments, one with teeth simulations and one with the wing of the fruit fly (Fig. 1). We compare three prediction methods: the multivariate breeder's equation, the method we introduce, and a univariate method based on realized heritability (h^2). This last method uses the time window of past generations to estimate the realized h^2 as the slope of the regression of the cumulative response to selection against the cumulative selection differential (11). This realized h^2 is then used to predict change at generation i.

The performance of the prediction methods is assessed by calculating the error for single-generation predictions obtained as the relative rms error (RMSE) between the multivariate series of predictions and the multivariate series of true changes. The relative RMSE is calculated as the squared root of the sum of the squared differences between true and predicted changes, relative to the true changes. As explained in *Materials and Methods*, the true change for the teeth experiments is directly the measured change. For the fly experiments, the true change is obtained from a quadratic regression of the means (Fig. 1*G*). The RMSE is a general measurement of the goodness of prediction for the whole time series, which is affected by both the accuracy (i.e., bias) and the precision (i.e., variance) of the predictors. It is, therefore, the main criteria to compare methods used in this work. Note that the RMSE reported here removes the error from the two first generations since at least two generations are required to estimate the first realized h^2.

Predicting the Response to Selection in Teeth Simulations. The teeth artificial selection experiments are in silico simulations of evolution in a population. A key feature of these simulations is that the mapping between genetic and phenotypic variation is done using a model of development that produces realistic morphological

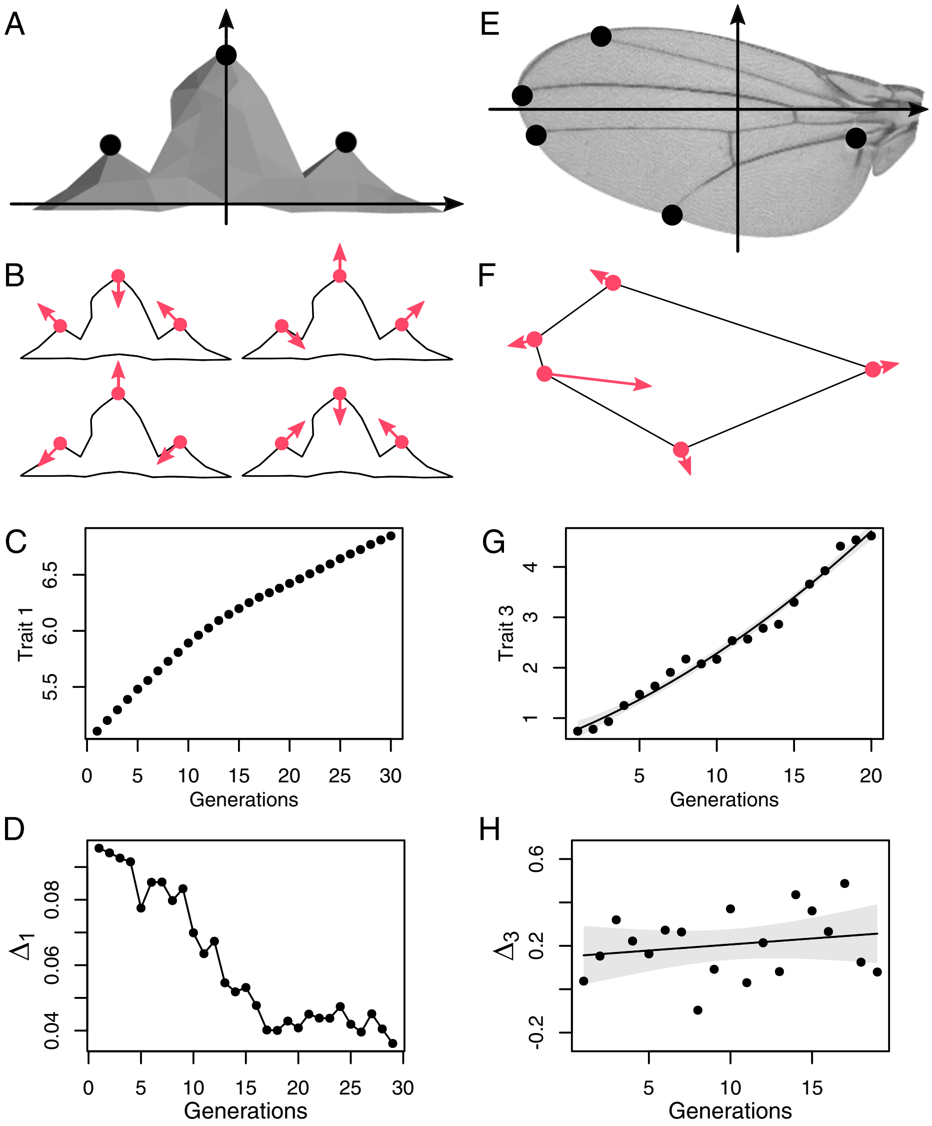

Fig. 1. Summary of the artificial selection experiments. *A–D* correspond to the teeth experiments, and *E–H* correspond to the fly wing experiments. *A* shows the tooth morphology and the three landmarks used for the experiments. The five traits are the *x* and *y* coordinates for anterior and posterior landmarks and the *y* coordinate for central landmark. *B* shows the directions of selection for 4 of the 32 evolutionary simulations as examples. *C* is the mean of trait 1 in time for simulation 2, and *D* shows the change in the trait mean. We do not make a regression in *C* because there is little measurement noise. *E* shows the morphology and the five landmarks on the wing. There are six phenotypic traits that are obtained after aligning the 10 coordinates of these landmarks using Procrustes superimposition. *F* shows the direction of selection. *G* shows the mean of trait 3 for line R1 together with a quadratic fit to the data and its 90% CI (remaining traits and replicate lines are shown in *SI Appendix*, Fig. S1). *H* shows the corresponding change in the mean of trait 3. The line with its 90% CI is the true change estimated from the quadratic regression shown in *G*, while the points are the measured changes with noise.

variation (34). Importantly, the genotype–phenotype map of this model is known to be complex and lead to biases in the estimation of the response to selection (26). There are a total of 32 simulations with different optima, each of 30 generations (Fig. 1*C*). Fig. 1 shows the tooth morphology and the three landmarks used. The x and y coordinates of these landmarks are the five measured traits. Fig. 1 also shows the response to selection for one trait in an example simulation.

Because the data are simulated, all conditions are controlled. This allows us to study the performance of the prediction methods in different scenarios, where specific sources of prediction error are isolated and when sources are interacting with each other. To study how using incorrect estimates of G_i and P_i affects predictions, we introduce the parameter m, which determines how often the estimates of P_i and G_i are updated (i.e., $m = 1$ means that P_i and G_i are updated in every generation). In the simulations, the selection differential s_i is known in each generation since all individuals are measured with no error. However, in many scenarios, there can be large uncertainty in the measurement of selection. This is the case when selection is not measured in every generation or when only a sample of the population is used to estimate selection, leading to SEs that can be as large as the measurement of selection itself (8, 35). We investigate the sensitivity of the different predictors to noise in the measurement of selection by introducing the parameter σ, which defines the distribution of noise that we add to s_i in each generation. For this, we multiply the selection differential s_i in each generation by $1 + u$, where u is a random number from a normal distribution with mean zero and SD σ. Finally, we also try different values of L, the number of past generations in the moving window used by the KF method (to learn the parameter ρ_i) and the realized h^2 method (to estimate the realized h^2). By trying different combinations of the above parameters, we can explore different scenarios and evaluate how well the methods predict. For example, for $m = 1$ and $\sigma = 0.3$, we can evaluate how the methods deal with noise when G_i and P_i are estimated in each generation, and for $m = 30$ and $\sigma = 0.3$, we can further study how the methods perform when the matrices G_i and P_i are assumed constant in addition to the stochastic noise in s_i.

Fig. 2*A* summarizes the total error for the different prediction methods in the teeth simulations. Different scenarios are explored, characterized by the combinations of m, σ, and L as described above. The figure shows that the KF method outperforms the alternatives on average. The KF method is particularly better than the breeder's equation when G_i and P_i are not updated in every generation (i.e., $m = 30$, regardless of σ and L) since it is able to correct the prediction bias generated by outdated estimates of variance components. In the idealized scenario with $m = 1$ and $\sigma = 0$, the breeder's equation is the least biased possible. In this case, the KF method reduces the spread of the prediction error toward lower values but cannot reduce the median error with respect to the breeder's equation. This is because the uncertainty of the method is sometimes larger than the value of the bias.

When compared with the realized h^2 method, the KF method is particularly better when selection is measured with noise ($\sigma = 0.3$) and when the size of the time window is larger ($L = 15$). This occurs because the KF method can manage noise more efficiently through the Kalman filter and because it is less sensitive to the size of the time window L. The latter results from the fact that the realized h^2 method assumes that the h^2 is constant inside the time window and therefore, has problems if it changes rapidly. The KF method, on the other hand, only assumes that the parameter ρ_i of the Kalman filter is constant inside the window, which is a more robust assumption. Further, the KF method greatly benefits from frequent updates of the variance components ($m = 1$), while the realized h^2 method is unable to exploit this information. An additional limitation of the realized h^2 method is that it is intrinsically univariate, so it cannot use information of selection on other traits.

Fig. 2*B* shows the predicted and true changes for some traits in example simulations for the scenario with $m = 30$, $\sigma = 0.3$, and $L = 5$ (the remaining traits are shown in *SI Appendix*, Fig. S2). Note that the difference between the prediction using the breeder's equation and the KF method in each generation i is the bias b_i, which is shown explicitly for each example in Fig. 2*C*.

We further use the teeth artificial selection experiments to study the situation where traits that are under selection are omitted from

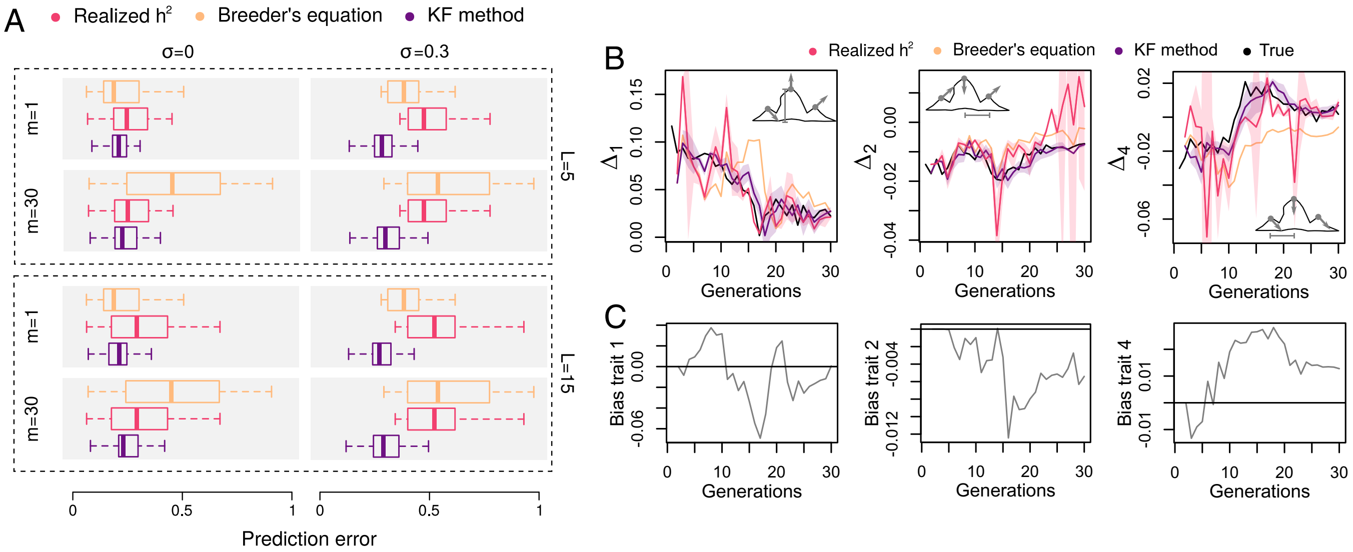

Fig. 2. Comparison of the prediction methods for the teeth experiments. *A* shows box plots of the prediction error for the 32 simulations using the tooth model for the three prediction methods compared (i.e., the KF method, the breeder's equation, and the realized h^2 method). Different scenarios were explored by modifying the parameters *L* (the size of the time window), *m* (the update time of variance component estimates), and σ (the stochastic error in the estimate of the selection differential). *B* shows examples of time series for the predictions and true changes with $m = 30$, $\sigma = 0.3$, and $L = 5$. The time series correspond to traits 1, 2, and 4 in simulations 6, 11, and 15, respectively (remaining traits are shown in *SI Appendix*, Fig. S2). The insets show a scheme of the tooth, with the direction of selection in each case represented with arrows and the trait plotted shown with a bar. *C* shows the dynamics of the bias for each of the examples given in *B*.

the prediction, what is known as the missing character problem. Specifically, we predict the change in traits 1, 2, and 3 without data from traits 4 and 5. As before, we explore different scenarios by varying the values of m, σ, and L. This allows us to further study how omitting traits interacts with other sources of bias, such as using outdated estimates of variance components (i.e., $m = 30$).

Fig. 3*A* shows a summary of the prediction errors of the different methods in the different scenarios studied. We find that omission of traits can lead to biases and that the KF method is able to correct the errors to a large extent. Fig. 3*B* includes the time series of the same trait in an example simulation, with predictions being done for different combinations of m, σ, and L. The corresponding bias is shown in Fig. 3*C*. Note that the bias becomes larger when in addition to the omission of traits in the prediction, it is assumed that variance components are constant (compare $m = 1$ and $m = 30$ in the time series). Further, note that the realized h^2 method becomes heavily biased for larger window size, while the KF method is robust to these changes (compare $L = 5$ and $L = 15$).

Predicting the Response to Selection in the Wing. The artificial selection experiments in the wing show the full complexity of the problem of predicting the response to selection in a real population. This is the most common scenario in which the KF method can be applied. There are three replicates with selection and one control line, all coming from the same base population and running for 20 generations (Fig. 1, *Materials and Methods*, and *SI Appendix*, Fig. S1). In each generation, 100 males and 100 females are measured. Selection is applied on five landmarks of the wing as shown in Fig. 1*E* by selecting the 50% of measured individuals in the direction shown in Fig. 1*F*.

For this experiment, we only calculate the G matrix at the beginning (i.e., G_1). Since the control line and the selection lines all start from the same base population, we use the pedigree and phenotypic data of the initial generations of the control to estimate G_1. We call the pedigree depth the number of generations of the control line used to estimate G_1. The larger the pedigree depth, the more precise the estimate of G_1. We test the predicting ability of the different prediction methods using estimates of G_1 for different pedigree depths ranging from 2 to 15, which correspond to 400 to 3,000 individuals from the control. We also include a pedigree depth of one, which means assuming that $G_1 = P_1$ (i.e., that all phenotypic variation is additive genetic).

Fig. 4 shows the predictions for the change in the traits using the KF method, the realized h^2 method, and the breeder's equation for the first three traits in selection line R2 (remaining traits and lines are shown in *SI Appendix*, Fig. S2). A pedigree depth of three was used here, and window sizes of 5 and 15 were explored. The time series of the predictions also includes the measured changes in each generation as well as the true change that is obtained from the quadratic regression of the response (*Materials and Methods*). It can be seen that the KF method yields predictions that are closer to the true change than the other methods on average.

We compare the predictions against the measured changes without the regression by calculating the cumulative error in generation i as the sum of the differences between the prediction and the measured change (i.e., without regression) from generation 2 to i. Cumulative errors are shown in Fig. 4 and confirm that the KF method has cumulative errors closer to zero. Further, note that the KF method is more robust to changes in the window size L than the realized h^2 method as seen also for the teeth data.

Fig. 5*A* shows the prediction error for the KF method, the realized h^2 method, and the breeder's equation for different pedigree depths and with a window size $L = 5$ ($L = 15$ gives very similar results as shown in *SI Appendix*, Fig. S4). For low pedigree depths (i.e., inaccurate estimates of G_1 and P_1), the realized h^2 method outperforms the breeder's equation, while the opposite is true for high pedigree depths. This is not surprising since the performance of the breeder's equation is largely affected by how well we can estimate the variance components, while the realized h^2 method is completely independent of the estimates of G_1 and P_1 and relies exclusively on past time series data. Note that our method exploits both sources of information, namely the time series and the estimates of the variance components.

Notably, the KF method using a G matrix with a small pedigree depth outperforms the breeder's equation using a G matrix with a large pedigree depth. This is important because experimentally, it

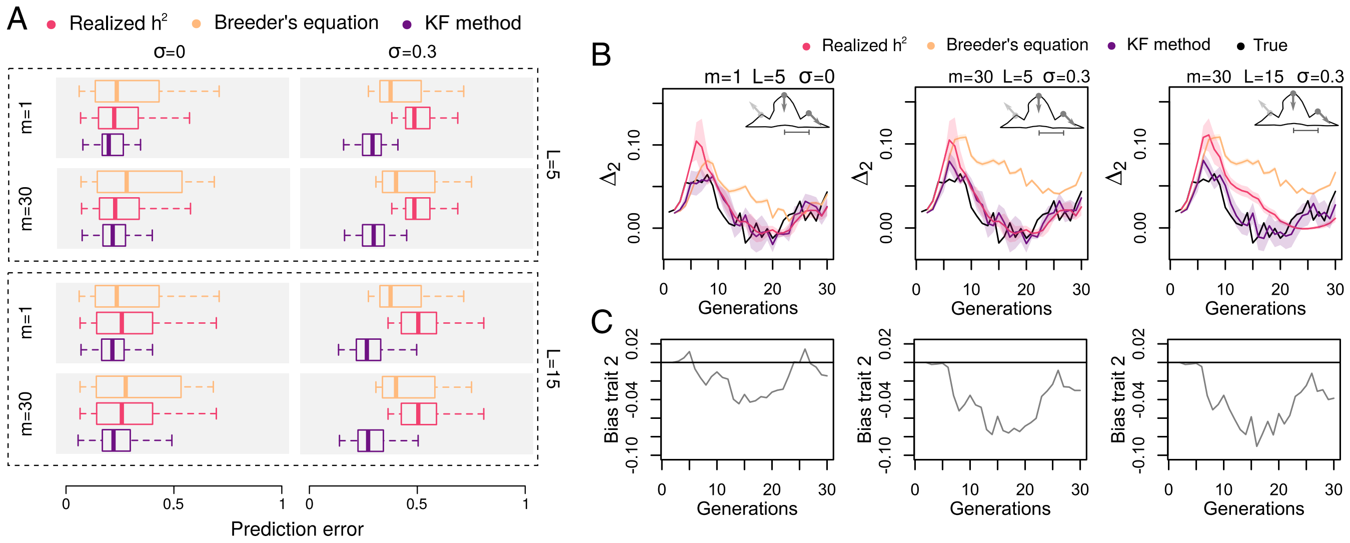

Fig. 3. Comparison of the prediction methods for the teeth experiments when traits are missing. *A* shows the prediction error for the 32 simulations using the tooth model when predictions for traits 1, 2, and 3 are made without information of traits 4 and 5. Errors are shown for different scenarios characterized by different combinations of parameters *L* (the size of the time window), *m* (the update time of variance component estimates), and σ (the stochastic error in the estimate of the selection differential). *B* shows the time series of predictions and true changes for trait 2 in example simulation 32. The insets show a scheme of the tooth, with the direction of selection in each case represented with arrows and the trait plotted shown with a bar. Different combinations of *m*, *L*, and σ were used in each plot. The corresponding dynamics of the prediction bias are shown in *C*.

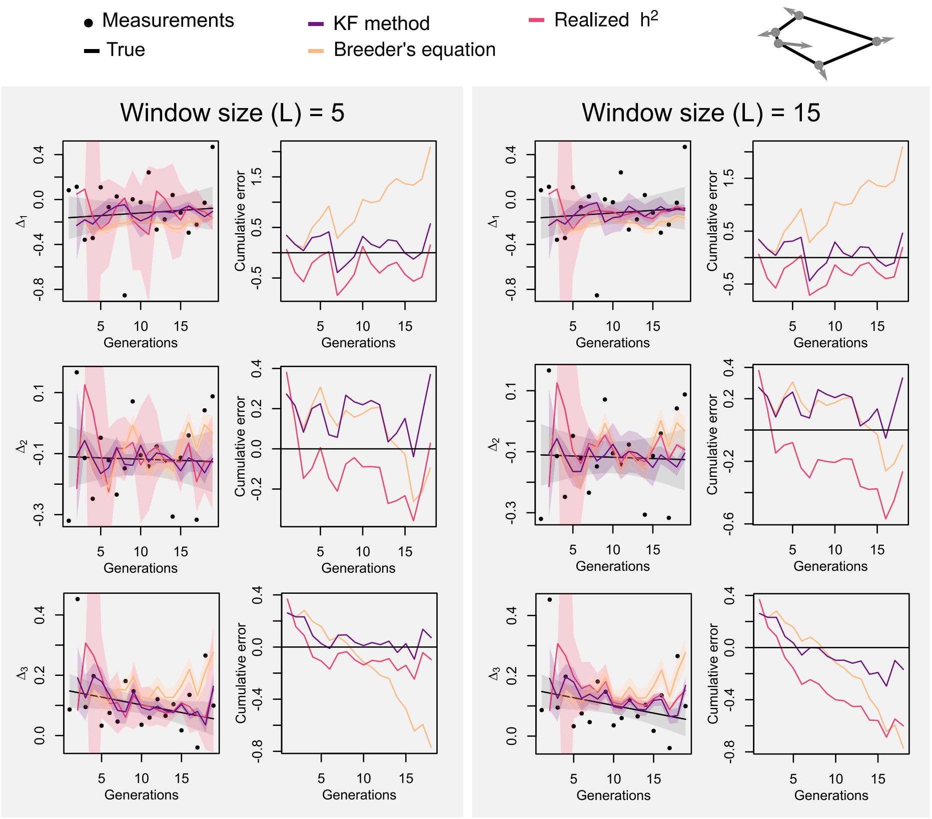

Fig. 4. Comparison of the prediction methods for the first three traits in selection line R2 of the fly wing experiments. The figure shows the time series for the different methods together with the corresponding cumulative errors for window sizes (*L*) of 5 and 15 generations in *Left* and *Right*, respectively. The time series of the predictions are shown on the left of each panel together with the true change (i.e., obtained from the regression of trait means) (*Materials and Methods*) and the measured change as points. The predicted and true changes are shown with their approximated uncertainties (90% CI for the true change, one SD for the predictions). The cumulative error for each method is calculated in each generation *i* as the sum of the difference between the prediction and the measured change from generation 2 to *i*. A pedigree depth of three was used to estimate the variance components. The rest of the traits for line R2 as well as the six traits for lines R1 and R3 are shown in *SI Appendix*, Fig. S3.

is much more expensive to increase the accuracy of the estimate of G_1 than to apply the KF method. The latter only requires recording the trait means in past generations, while the former requires phenotypic and relatedness data in particular breeding designs.

We found that there are two sources of prediction bias for the breeder's equation in these experiments. First, there is bias associated with using wrong estimates of G_1 and P_1. This prediction error is most evident when using estimates of G_1 with low pedigree depth and even more when assuming $G_1 = P_1$. Increasing the pedigree depth of the estimates can correct much of these error. The second source of bias is the fact that G_i changes during the experiment. This leads to possible errors at later stages of the experiment if the G matrix estimated for the base population is used, even if the estimate of G_1 is obtained with high accuracy (i.e., a deep pedigree in this case).

Both of the errors described above can be seen in Fig. 5*B* for trait 6 of line R1 using different pedigree depths, namely 1 ($G = P$), 4, and 10. The corresponding time series of the bias is shown in Fig. 5*C*. A big part of the error is reduced when increasing the pedigree depth. However, even when using a precise estimate of G_1 (pedigree depth of 10), the breeder's predictions remains biased toward the end of the experiment.

Discussion

We developed a method to predict the response to directional selection by combining the breeder's equation with data from the time series. We tested this method, which we refer to as the KF method, with two sets of artificial selection experiments and show that it outperforms the multivariate breeder's equation and a univariate method based on realized heritability on average. The method is general and can be applied to virtually any evolving system that is under sustained directional selection. Most importantly, the KF method only requires the record of means of the trait for past generations, which is relatively easy to collect, at least compared with alternatives like obtaining better estimates of G_i. In this way, the method can be applied to a wide variety of scenarios, especially when the assumptions of the breeder's equation are not met, like in the later stages of long-term selection studies when the full set of traits under selection is not known and when G_i cannot be accurately estimated. The more the assumptions are violated, the more the KF method will outperform the breeder's equation, as shown in Figs. 2, 3, and 5.

We discuss three key elements of the KF method. The first key element is the introduction of the bias term in Eq. **2**. This is proposed on the grounds of previous theoretical and empirical work that shows that the expected value of the prediction error using the breeder's equation may not be zero (8, 14, 15, 25). The bias is introduced here as a single term, b, which can be understood as the quantitative effect of violating the assumptions of the breeder's equation, and its value reflects the complex mixture of sources that is specific to each system. This simple way of modeling the bias allows for improved predictions but makes it hard to disentangle how different sources of prediction error quantitatively contribute to the total value of the bias. If this separation of the bias is desirable, however, it could be performed after the experiment is over and the complete time series of the bias is available. For example, it is possible to perform analyses of the type proposed

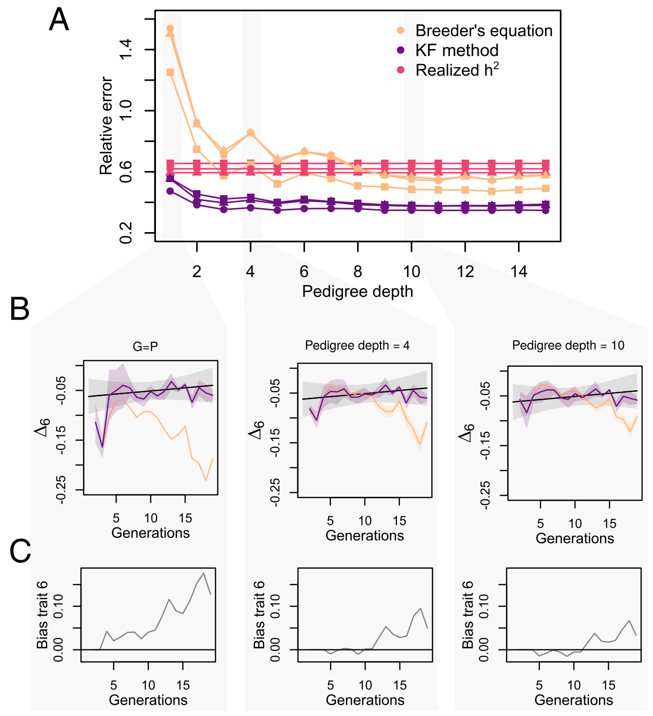

Fig. 5. Comparison of prediction methods for different pedigree depths in the artificial selection experiments of the wing. *A* shows the relative RMSE of the prediction using the breeder's equation, the method based on realized h^2, and the KF method for the three replicated selection lines (squares correspond to line R1, circles correspond to line R2, and triangles correspond to line R3). The G and P matrices were estimated in the first generation with varying pedigree depths. A pedigree depth of one means that it is assumed that $G = P$. *B* shows the predicted and true time series for trait 6 of R1 for different pedigree depths. The corresponding bias is shown in *C*.

by Le Rouzic et al. (36) to analyze time series data, which are based on proposing autoregressive models for the parameters and keeping the model that maximizes the likelihood of the data.

The second key element of the method is the use of a Kalman filter. This was necessary to minimize the effect of stochastic noise on the estimation of the bias, which can be large (11, 30). Figs. 2 and 3 show that the Kalman filter is very robust to stochastic errors, particularly when compared with the alternative methods.

The third key element of the KF method is that we use a window of data to learn the parameter ρ_i of the filter in each generation using a learning algorithm. Apart from enabling the method to be used in real time (i.e., during the experiment), it has the important quality that it allows the parameters to change in time. Moreover, it exploits the dynamical properties of the time series (36), which are specific to the population of interest and its singularities. Note that the moving time window is used to learn the parameter ρ_i, which is then used by the filter to estimate the bias in each generation. In this way, the time window is not used to estimate the value of the bias directly as the average of the prediction errors using the breeder's equation in the window. This is a valid way of estimating the bias, but it provides worse estimates than the method based on the Kalman filter (over 23% worse for the fly wing experiments) (*SI Appendix*, Table S1). Note that here we propose to reduce the covariance matrices R_i and Q_i to the single parameter ρ_i, allowing for a straightforward learning algorithm that relies on exploring different values of ρ inside the sliding time window. This simple method to estimate the covariance matrices gives very good predictions in the datasets we explore, but other more complex methods can be used for this purpose (31–33).

The method introduced here can be classified as recursive because it forecasts the variables of interest in i using the estimate in $i-1$. There has been recent interest in recursive models to make predictions of future evolution (1, 36–39). For example, Nosil et al. (1) fitted an autoregressive model using several years of data of frequency changes of coloration and pattern in a population of stick insects. They examined whether data from early time points in the series could predict data in later time points of the series (similar to what we do using the window of past generations). They were able to successfully predict changes in frequency for a trait under clear frequency-dependent selection but failed to predict change for a trait under a more complex, unknown form of selection. They conclude that predictability was limited by the understanding of selection. The authors suggest that knowledge of selection could be determinant in improving predictions when using recursive models. The method we propose in this paper does exactly this; it combines a recursive model with knowledge of selection given by the breeder's equation. Used like this, the breeder's prediction contributes the type of information that purely recursive models are lacking. At the same time, purely theoretical models, like the breeder's equation, are based on simplifying assumptions that may miss some of the complexity of the system. The efficient combination of the recursive model, which is data driven, and the breeder's equation, which is theoretical, is what results in the method proposed here to outperform each approach when used separately.

A notable finding is that the KF method provides good predictions even when G is not estimated at all. This is assuming that $G = P$. This is shown in Fig. 5 for the fly experiments and *SI Appendix*, Fig. S5 for the teeth experiments. This is important because the P matrix has been used as a proxy of the G matrix for morphological traits, a simplification suggested due to the difficulty in estimating the latter. The simplification that G is proportional to P is known as Cheverud's conjecture (40–42) and has been used, for example, in ref. 43 to infer the evolution of patterns of genetic covariation under directional selection. When used in this method, assuming $G = P$ still provides good predictions because the resulting deviations are corrected by the bias term. Note that even in this case, the information of selection is still exploited, as it enters the predictions through the selection differential, $\boldsymbol{s}_i$. An important note is that both the breeder's equation and the KF method perform significantly better when G_i is estimated than when it is assumed that $G_i = P_i$ (compare precision 1 and 2 in Fig. 5*A*). This means that G_i contains useful information, even when estimated with relatively low precision. This improvement does not occur for the realized h^2 method, which is unable to incorporate Information from variance components and only relies on time series data.

The method proposed here was developed under the assumption of directional selection sustained for several generations. The method is, therefore, limited to this type of selection. However, the formalism of the Kalman filter has the potential to be used for evolutionary predictions in several other scenarios, particularly when stochastic noise is a serious issue. For example, the Kalman filter could be coupled with an autoregressive model to predict evolution under frequency-dependent selection based on previous evidence that such models are adequate for this type of selection (1). For the case of fluctuating selection, prediction is severely limited by the difficulty in obtaining information of how selection is acting in each generation. Recent efforts (39) have tried to map environmental fluctuations to fluctuations in selection since certain environmental cues, such as temperature, are much easier to measure than selection itself. By modifying the equations that relate the states with the measurements, one could include the information of these other environmental cues to improve the predictions under the Kalman filter formalism similar to the one introduced here.

Data-driven methods are only becoming more popular in the future. This change from more classical, theoretical methods is fueled by the rapid accumulation of data. The method we propose here is line with this change by combining theory and data. As suggested by other authors (38), this is a promising future for developing better predictions in evolutionary biology.

Materials and Methods

Experiment 1: Teeth. We used data of in silico artificial selection experiments on teeth. Details of the simulations are given in previous work, and the data are publicly available (19, 26). Briefly, each evolutionary simulation has a population of genotypes. Each genotype is mapped to a tooth morphology through a deterministic model of tooth development (34, 44). The tooth model recapitulates the process of development for a tooth, starting from a flat epithelium to a complex three-dimensional morphology. The dynamics of development are determined by the value of a set of parameters that are determined by the genotype. Traits were measured on each tooth. These were the *x* and *y* coordinates of three landmarks located in the three tallest cusps of the tooth (Fig. 1*A*). In each generation, once the genotypes of all individuals had been mapped to their corresponding phenotypes using the tooth development model, selection was applied by choosing 50% of the individuals with morphology closest to the optimum. Each simulation had an optimum shape defined at the beginning, which determined the direction of selection (Fig. 1*C*). Selected parents were paired randomly and produced the next generation of genotypes. Each couple produced four offspring, resulting in a constant population size. Recombination and mutation were included in each generation, and the process was iterated to simulate evolution. There are in total 32 simulations, each with a different selection optimum. Each simulation was run for 30 generations using a population of 300 males and 300 females.

Estimation of Variance Components and True Change. In each generation, the elements of the breeder's equation were estimated (i.e., G_i, P_i, and $\boldsymbol{s}_i$). Variance components were estimated from a half-sibling breeding design using individuals at generation *i* as the base population (details are in ref. 25). The animal model used was the simplest possible (i.e., with only additive genetic merit fitted to each individual). Restricted maximum likelihood (REML) estimates of G_i and P_i were obtained using the software WOMBAT (45). Sampling variation in the estimation of G_i was accounted for using the REML–multivariate normal (REML-MVN) method (46). For each generation, we resampled 100 G and P matrices from this distribution and used them to calculate 100 predicted changes using the selection differential and the breeder's equation. We plot the mean and one SD of these predictions. Note that the tooth development model is deterministic and that there is no measurement error. Moreover, we have a large sample size. This allows for very precise estimates of G_i and P_i. Due to the fact that there is little measurement noise for the population mean in the simulations, the true change was obtained directly as $\Delta_i = \bar{z}_{i+1} - \bar{z}_i$. This is the amount that we look to predict at generation *i* (Fig. 1 *E* and *G*).

Experiment 2: Fruit Fly Wing. We performed artificial selection experiments on the wing of the fruit fly *Drosophila melanogaster*. The starting population was founded from 250 isofemale lines derived from flies captured during the summer of 2017 in Groningen, the Netherlands by the Billeter laboratory. From each line, 25 females and males were collected and merged to make a large outbred population that was maintained in laboratory conditions. For the initial generation of the experiments, 100 virgin males and 100 virgin females from the large population were randomly assigned to one of four lines. Three of these lines were subjected to selection (R1, R2, and R3), with the remaining being a control without selection (C1). Lines were kept at 25 °C with alternating 12-h light and dark cycles during the experiment.

In each generation, 100 males and 100 females were collected as virgins. The left wing of each collected, anesthetized fly was taken by the automatic system known as the WingMachine (47, 48). The *x* and *y* coordinates of the five landmarks shown in Fig. 1*B* were obtained using a semiautomatic landmarking software (47). In the control line, 50 males and 50 females were chosen randomly as parents for the next generation. In the selected lines, the 50 males and 50 females with wings with the shortest distance to the optimum morphology were selected as parents. The distance of each individual to the optimum was calculated as the Euclidean distance between the values of the traits in the individuals and the optimal values of the traits. The optimum morphology is shown in Fig. 1*D*, and it is the same for the three lines with selection. The process of image processing and selection was repeated in each generation. Sibling mating was avoided to reduce inbreeding. The process was repeated for a total of 20 generations, equivalent to 4,000 flies per line (16,000 in total). If some of the formed couples did not produce offspring for the next generation either because one of the parents died or due to infertility, we measured more offspring from other couples to complete the 200 individuals per generation. We also formed three extra couples in each generation to provide extra individuals in case some of the original 50 couples failed to produce offspring.

As mentioned above, we measured the *x* and *y* coordinates of five landmarks, resulting in 10 traits. The data were aligned by generalized Procrustes least squares superimposition. Four degrees of freedom are lost in this process: one to estimate wing size and three to standardize the orientation of wing shapes. Therefore, there are only six independent traits in the data. For these traits to be comparable between lines and through the generations, we use the six first components of a principal component analysis of generation 1 of the control as a reference and project all the data to that space. The resulting six phenotypic traits are a linear combination of the original 10 traits that conserves all relevant variation in all lines. In this paper, we refer to these six traits as the phenotypic traits. The means of these traits against generations for all four experimental lines are shown in *SI Appendix*, Fig. S1.

Estimation of variance components and true change. All lines start from the same founding population. We estimate G_1 and P_1 for this founding population by pooling the first *n* generations of the control. For these *n* generations, we have the pedigree and phenotypic data. We call *n* the depth of the pedigree. Here, we explore values of *n* from 2 to 15. REML estimates of G_1 and P_1 were obtained using the software WOMBAT (45), and sampling variation was estimated using the REML–MVN method (46). The linear, mixed effect model (i.e. animal model) used included sex, generation, and identification of the person measuring as fixed effects.

The estimation of the means in each generation inevitably has noise. Noise arises from the imaging and landmarking process, finite sampling of the population, and drift. Because we focus on directional selection, this noise has to be removed. We perform a quadratic regression to the 20-generation time series of the means, which is the standard type of regression used for long-term artificial selection data (8, 49–51). The fitted values are used as Δ_i, which we call the true change. This is compared with the change predicted by the different methods to calculate the prediction error (Fig. 1 *F* and *H*). We also compare predictions with the measured change in trait means (i.e., without regression) by calculating the cumulative prediction error. For generation *i*, the cumulative error is the sum of the differences between the measured change and the predicted change from generation 1 to *i*.

Appendix A

Here, we will derive the equations of the Kalman filter and show how the gain matrix K_i is calculated in each generation. For this, we first express the model in matrix notation. The state equation is given by

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} + \boldsymbol{\eta}_i, \qquad [5]$$

where $\boldsymbol{x}_i = (\Delta_{i-1}, b_i)^T$ and $\boldsymbol{\eta}_i = \left(\eta_i, \eta_i^b\right)^T$. The measurements used in the model are the prediction of the breeder's equation, $\tilde{\Delta}_i^B$, and the most recent measured change in trait means, $\tilde{\Delta}_{i-1}$. Δ_i^B is related to the states via the following relationship:

$$\Delta_i^B = \Delta_i - b_i = \Delta_{i-1} + \eta_{i+1} - b_i. \qquad [6]$$

The measurement equation is then

$$\boldsymbol{y}_i = C\boldsymbol{x}_i + B\boldsymbol{\eta}_{i+1} + \boldsymbol{v}_i \qquad [7]$$

with $\boldsymbol{y}_i = \left(\tilde{\Delta}_i^B, \tilde{\Delta}_{i-1}\right)^T$, $\boldsymbol{v}_i = \left(v_i^B, v_i\right)^T$, and

$$C = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}; \; B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}; \qquad [8]$$

the state-space model is determined by [5] and [7]. The objective of the Kalman filter is to provide estimates of the states at generation i ($\hat{\boldsymbol{x}}_i$) using the previous estimate ($\hat{\boldsymbol{x}}_{i-1}$) and new measurements ($\boldsymbol{y}_i$). Particularly, the Kalman filter provides the $\hat{\boldsymbol{x}}_i$ that minimizes the variance of the error $\boldsymbol{e}_i = \boldsymbol{x}_i - \hat{\boldsymbol{x}}_i$ denoted by $\Phi_i = \mathcal{E}[\boldsymbol{e}_i\boldsymbol{e}_i^T]$. For this, it uses the expression

$$\hat{\boldsymbol{x}}_i = \hat{\boldsymbol{x}}_{i-1} + K_i(\boldsymbol{y}_i - C\hat{\boldsymbol{x}}_{i-1}), \qquad [9]$$

which depends on the matrix K_i. To find the K_i that minimizes Φ_i, we first need to obtain an expression for the error. Using [5], [7], and [9], we get

$$\begin{aligned} \boldsymbol{e}_i &= \boldsymbol{x}_i - \hat{\boldsymbol{x}}_i \\ &= \begin{pmatrix} I & -K_i \end{pmatrix} \left[\begin{pmatrix} I \\ C \end{pmatrix} \boldsymbol{e}_{i-1} + \begin{pmatrix} \boldsymbol{\eta}_i \\ C\boldsymbol{\eta}_i + B\boldsymbol{\eta}_{i+1} + \boldsymbol{v}_i \end{pmatrix} \right], \end{aligned} \qquad [10]$$

where I is the identity matrix. Considering η_i, v_i and e_{i-1} independent, the expected value of the cross-products between $\boldsymbol{e}_{i-1}$ and both $\boldsymbol{\eta}_i$ and $\boldsymbol{v}_i$ vanishes. Then, we get

$$\begin{aligned} \Phi_i = \begin{pmatrix} I & -K_i \end{pmatrix} &\left[\begin{pmatrix} I \\ C \end{pmatrix} \Phi_{i-1} \begin{pmatrix} I \\ C \end{pmatrix}^T \right. \\ &\left. + \begin{pmatrix} Q_i & Q_i C^T \\ CQ_i & CQ_i C^T + R_i^* \end{pmatrix} \right] \begin{pmatrix} I \\ -K_i^T \end{pmatrix}, \end{aligned} \qquad [11]$$

where $R_i^* = BQ_{i+1}B^T + R_i$, $Q_i = \mathcal{E}[\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T]$, and $R_i = \mathcal{E}[\boldsymbol{v}_i\boldsymbol{v}_i^T]$. We want to find K_i such that Φ_i is minimized. This is a convex quadratic minimization problem with a unique solution that can be obtained, for example, by using the method of completing the squares (ref. 29, pp. 430–433). The solution is given by

$$\begin{aligned} K_i &= (\Phi_{i-1} + Q_i)C^T(C(\Phi_{i-1} + Q_i)C^T + R_i^*)^{-1}, && [12] \\ \hat{\boldsymbol{x}}_i &= \hat{\boldsymbol{x}}_{i-1} + K_i(\boldsymbol{y}_i - C\hat{\boldsymbol{x}}_{i-1}), && [13] \\ \Phi_i &= (I - K_i C)(\Phi_{i-1} + Q_i). && [14] \end{aligned}$$

From this recursion, we get estimates of the states $\hat{\Delta}_{i-1}$ and $\hat{b}_i$. Finally, the prediction at generation i is given by

$$\hat{\Delta}_i = \tilde{\Delta}_i^B + \hat{b}_i. \qquad [15]$$

Appendix B

To implement the Kalman filter given in *Appendix A*, we need the matrices Q_i and R_i^*. These parameters are, however, generally unknown and have to be identified from the data. Here, we use a simple method based on exploring different values of the parameters and keeping the ones that result in the smallest prediction error in a moving time window. For this, we reduce the matrices to a single parameter ρ_i by assuming that the matrices are diagonal with equal elements in the diagonal: that is, $Q_i = q_i\, I$ and $R_i^* = r_i^*\, I$, where I is the 2×2 identity matrix. The reduction is justified by the fact that the noise in the measurements, as well as in the states, is in the same units and of the same order of magnitude. We further assume that these values are constant inside the window: that is, $q_i \approx q_{i+1}$ and $r_i^* \approx r_{i+1}^*$. If we use these definitions, we can rewrite the equations from *Appendix A* as

$$\begin{aligned} K_i &= (\Phi_{i-1}^* + \rho_i I)C^T(C(\Phi_{i-1}^* + \rho_i I)C^T + I)^{-1}, && [16] \\ \hat{\boldsymbol{x}}_i &= \hat{\boldsymbol{x}}_{i-1} + K_i(\boldsymbol{y}_i - C\hat{\boldsymbol{x}}_{i-1}), && [17] \\ \Phi_i^* &= (I - K_i C)(\Phi_{i-1}^* + \rho_i I), && [18] \end{aligned}$$

where we define $\Phi_i^* = \Phi_i / r_i^*$ and $\rho_i = q_i / r_i^*$. For each generation i and for a window size of L, the Kalman filter is run for generations $k \in \{i - L, \ldots, i - 1\}$ with different values of ρ_i and using Eqs. **16–18**. Then, ρ_i is chosen as the value that minimizes the mean square error between the predictions $\hat{\Delta}_k$ and the true value Δ_k. For the teeth simulations, Δ_k is directly the measured change in trait mean (i.e., $\tilde{\Delta}_k$) since there is very little measurement noise for the trait means. For the wing data, a better estimate of Δ_k is obtained by first making a linear regression of the means of the traits inside the window and then calculating the change in trait mean as the slope of the regression [**11**]. For the first two iterations of the algorithm, the window is too small to calculate ρ_i, so we set it to zero. The initial conditions are $\hat{\boldsymbol{x}}_1 = (\tilde{\Delta}_1^B, 0)^T$ and $\Phi_1^* = \mathbf{0}$.

Data Availability. The data from in-silico experiments using the tooth model is available in Dryad at https://doi.org/10.5061/dryad.9cnp5hqdr (52). The R script to apply the method and the data from the *D. melanogaster* experiments are stored in GitHub at https://github.com/millisan/Learning-from-mistakes (53).

ACKNOWLEDGMENTS. We thank Ruben H. Milocco for advice on applying the Kalman filter, David Houle and his team for instructions on how to perform artificial selection experiments with the fruit fly using the WingMachine, Pınar Güler Kohlmeier from the Billeter team for providing the flies from the wild population, and Jukka Jernvall for useful comments. We also thank Niko Björkell, Shuja Wahid, Ardit Fejzullahi, and Sara Goncalves for help carrying out the experiments. The research was funded by Academy of Finland Grant 315740, the Integrative Life Sciences Graduate Program of the University of Helsinki, Finnish Cultural Foundation Grant 00210730, Spanish Ministry of Science and Innovation Grant PGC2018-096802-B-I00, and the Spanish State Research Agency through the Severo Ochoa and María de Maeztu Program for Centers and Units of Excellence. We also thank Centres de Recerca de Catalunya Programme/Generalitat de Catalunya for institutional support. *Drosophila* work was supported by the Hi-Fly core facility funded by the Helsinki Institute of Life Science and Biocenter Finland.

1. P. Nosil *et al.*, Natural selection and the predictability of evolution in *Timema* stick insects. *Science* **359**, 765–770 (2018).
2. R. G. Shaw, From the past to the future: Considering the value and limits of evolutionary prediction. *Am. Nat.* **193**, 1–10 (2019).
3. A. Le Rouzic *et al.*, Unidirectional response to bidirectional selection on body size II. Quantitative genetics. *Ecol. Evol.* **10**, 11453–11466 (2020).
4. M. Wortel *et al.*, The why, what and how of predicting evolution across biology: From disease to biotechnology to biodiversity. EcoEvoRxiv [Preprint] (2021). https://ecoevorxiv.org/4u3mg/ (Accessed 28 September 2021).
5. R. Gomulkiewicz, R. G. Shaw, Evolutionary rescue beyond the models. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120093 (2013).
6. S. Cobey, Modeling infectious disease dynamics. *Science* **368**, 713–714 (2020).
7. D. A. Roff, A centennial celebration for quantitative genetics. *Evolution* **61**, 1017–1032 (2007).
8. B. Walsh, M. Lynch, *Evolution and Selection of Quantitative Traits* (Oxford University Press, 2018).
9. J. L. Lush, *Animal Breeding Plans*. (Iowa State College Press, ed. 2, 1937).
10. R. Lande, Quantitative genetic analysis of multivariate evolution, applied to brain: Body size allometry. *Evolution* **33**, 402–416 (1979).
11. M. Lynch, B. Walsh, *Genetics and Analysis of Quantitative Traits* (Oxford University Press, 1998).

12. L. E. B. Kruuk, Estimating genetic parameters in natural populations using the "animal model." *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**, 873–890 (2004).
13. R. Lande, S. J. Arnold, The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226 (1983).
14. S. H. Rice, *Evolutionary Theory: Mathematical and Conceptual Foundations* (Sinauer Associates, 2004).
15. B. Pujol *et al.*, The missing response to selection in the wild. *Trends Ecol. Evol.* **33**, 337–346 (2018).
16. S. J. Steppan, P. C. Phillips, D. Houle, Comparative quantitative genetics: Evolution of the G matrix. *Trends Ecol. Evol.* **17**, 320–327 (2002).
17. J. D. Aguirre, E. Hine, K. McGuigan, M. W. Blows, Comparing G: Multivariate analysis of genetic variation in multiple populations. *Heredity* **112**, 21–29 (2013).
18. A. G. Jones, S. J. Arnold, R. Bürger, Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution* **57**, 1747–1760 (2003).
19. L. Milocco, I. Salazar-Ciudad, The evolution of the G-matrix under nonlinear genotype-phenotype maps. *Am. Nat.* **199**, 420–435 (2022).
20. C. M. Sgrò, A. A. Hoffmann, Genetic correlations, tradeoffs and environmental variation. *Heredity* **93**, 241–248 (2004).
21. C. W. Wood, E. D. Brodie III, Environmental effects on the structure of the G-matrix. *Evolution* **69**, 2927–2940 (2015).
22. M. Pigliucci, Genetic variance–covariance matrices: A critique of the evolutionary quantitative genetics research program. *Biol. Philos.* **21**, 1–23 (2006).
23. A. Gimelfarb, J. H. Willis, Linearity versus nonlinearity of offspring-parent regression: An experimental study of *Drosophila melanogaster*. *Genetics* **138**, 343–352 (1994).
24. M. B. Morrissey, L. E. B. Kruuk, A. J. Wilson, The danger of applying the breeder's equation in observational studies of natural populations. *J. Evol. Biol.* **23**, 2277–2288 (2010).
25. L. Milocco, I. Salazar-Ciudad, Is evolution predictable? Quantitative genetics under complex genotype-phenotype maps. *Evolution* **74**, 230–244 (2020).
26. J. Merilä, B. C. Sheldon, L. E. B. Kruuk, Explaining stasis: Microevolutionary studies in natural populations. *Genetica* **112-113**, 199–222 (2001).
27. J. S. Heywood, An exact form of the breeder's equation for the evolution of a quantitative trait under natural selection. *Evolution* **59**, 2287–2298 (2005).
28. R. E. Kalman, A new approach to linear filtering and prediction problems. *J. Fluids Eng.* **82**, 35–45 (1960).
29. K. Åström, B. Wittenmark, *Computer-Controlled Systems: Theory and Design* (Prentice-Hall, Englewood Cliffs, NJ, ed. 3, 1997).
30. C. Pélabon *et al.*, Quantitative assessment of observed versus predicted responses to selection. *Evolution* **75**, 2217–2236 (2021).
31. M. Laine, "Introduction to dynamic linear models for time series analysis" in *Geodetic Time Series Analysis in Earth Sciences*, J.-P. Montillet, M.S. Bos, Eds. (Springer, Cham, Switzerland, 2020), pp. 139–156.
32. J. Duník, O. Straka, O. Kost, J. Havlík, Noise covariance matrices in state-space models: A survey and comparison of estimation methods—part I. *Int. J. Adapt. Control Signal Process.* **31**, 1505–1543 (2017).
33. Y. Huang, F. Zhu, G. Jia, Y. Zhang, A slide window variational adaptive Kalman filter. *IEEE Trans. Circuits Syst. II Express Briefs* **67**, 3552–3556 (2020).
34. I. Salazar-Ciudad, J. Jernvall, A computational model of teeth and the developmental origins of morphological variation. *Nature* **464**, 583–586 (2010).
35. M. B. Morrissey, In search of the best methods for multivariate selection analysis. *Methods Ecol. Evol.* **5**, 1095–1109 (2014).
36. A. Le Rouzic, D. Houle, T. F. Hansen, A modelling framework for the analysis of artificial-selection time series. *Genet. Res.* **93**, 155–173 (2011).
37. M. Rescan, D. Grulois, E. Ortega-Aboud, L. M. Chevin, Phenotypic memory drives population growth and extinction risk in a noisy environment. *Nat. Ecol. Evol.* **4**, 193–201 (2020).
38. P. Nosil *et al.*, Ecology shapes epistasis in a genotype-phenotype-fitness map for stick insect colour. *Nat. Ecol. Evol.* **4**, 1673–1684 (2020).
39. M. Rescan, D. Grulois, E. O. Aboud, P. de Villemereuil, L. M. Chevin, Predicting population genetic change in an autocorrelated random environment: Insights from a large automated experiment. *PLoS Genet.* **17**, e1009611 (2021).
40. J. M. Cheverud, A comparison of genetic and phenotypic correlations. *Evolution* **42**, 958–968 (1988).
41. S. M. Sodini, K. E. Kemper, N. R. Wray, M. Trzaskowski, Comparison of genotypic and phenotypic correlations: Cheverud's conjecture in humans. *Genetics* **209**, 941–948 (2018).
42. A. C. Love *et al.*, Evolvability in the fossil record. Paleobiology, 1-24 (9 November 2021). https://doi.org/10.1017/pab.2021.36.
43. A. P. A. Assis, J. L. Patton, A. Hubbe, G. Marroig, Directional selection effects on patterns of phenotypic (co)variation in wild populations. *Proc. R. Soc. B Biol. Sci.* **283**, 20161615 (2016).
44. E. Harjunmaa *et al.*, Replaying evolutionary transitions from the dental fossil record. *Nature* **512**, 44–48 (2014).
45. K. Meyer, WOMBAT: A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J. Zhejiang Univ. Sci. B* **8**, 815–821 (2007).
46. D. Houle, K. Meyer, Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood. *J. Evol. Biol.* **28**, 1542–1549 (2015).
47. D. Houle, J. Mezey, P. Galpern, A. Carter, Automated measurement of *Drosophila* wings. *BMC Evol. Biol.* **3**, 25 (2003).
48. J. G. Mezey, D. Houle, The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. *Evolution* **59**, 1027–1038 (2005).
49. E. J. Eisen, Long-term selection response for 12-day litter weight in mice. *Genetics* **72**, 129–142 (1972).
50. J. J. Rutledge, E. J. Eisen, J. E. Legates, An experimental evaluation of genetic correlation. *Genetics* **75**, 709–726 (1973).
51. P. Grassini, K. M. Eskridge, K. G. Cassman, Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nat. Commun.* **4**, 2918 (2013).
52. L. Milocco, Data from "Is evolution predictable? Quantitative genetics under complex genotype-phenotype maps. Dryad. https://doi.org/10.5061/dryad.9cnp5hqdr. Deposited 18 December 2019.
53. L. Milocco, millisan/Learning from mistakes. GitHub. https://github.com/millisan/Learning-from-mistakes. Deposited 7 February 2022.