**STEP-BY-STEP FULL MODULE**

The last module of SQuID allows the user to generate data sets of individuals repeatedly expressing phenotypes, for one or two traits, in uniform time. The user has thus the flexibility to add different variance components that will form the phenotype of the individual at each time step, and to set up a relative importance to each component. SQuID also allows the user to collect a sample of the simulated phenotypes (i.e. the operational data set), according to specific sampling design. Finally, the user has the opportunity to download the operational data set for further analyses.

In this page, we will introduce to you step by step how to use our full module. In case you are already familiar with the full model we invite to switch to the full model express page, which is more straightforward.

**DESCRIPTION TAB**

In this section, we will describe the model used to simulate individual phenotype values and the model used to create the sampling design.

**SIMULATION MODEL**

As a first step, we generate phenotype values of individuals that belong to a study population (i.e. replicate world). Phenotypic values of each trait are calculated using the phenotypic equation described below.

**Model equation**

*yhij* = (*0y* + *Iyi*) + (*1y* + *S1yi*).*x1hij* + (*2y* + *S2yi*).*x2hij* + (*12y* + *S12yi*)*x1hij*.*x2hij* + *Gyj* + *eyhij*

*zhij* = (*0z* + *Izi*) + (*1z* + *S1zi*).*x1hij* + (*2z* + *S2zi*).*x2hij* + (*12z* + *S12zi*)*x1hij*.*x2hij* + *Gzj* + *ezhij*

|  |  |
| --- | --- |
| Components | Explanations |
|  | instance of time |
|  | individual identifier |
|  | group identifier |
|  | phenotypic value of trait y at instance h for individual i within group j |
|  | phenotypic value of trait z at instance h for individual i within group j |
|  | population mean |
|  | individual-specific deviations (random-intercepts) for individual i |
|  | population-wide response to environmental influences x |
|  | individual-specific response to environmental influences x (random-slope) for individual i |
|  | Higher-level grouping for each group j |
|  | residual for the individual i and the group j at time h |

The equation represents the model that generated phenotypic values (*yhij* or *zhij*) of two traits for each individual *i* and at each instance of time *h*, in each group *j*.

It should be noted that subscripts are organised to represent the different hierarchical levels involved in the generation of the value of the trait: subscript *h* stands for the instance at which the traits is measured for the individual *i*, within group *j.* Finally subscript *y* or *z* refers to the two traits generated.

At first sight we can see that the phenotypic value of the trait *y* for individual *i* at instance *h* in group *j* is the sum of a series of components. Let’s first decompose the equation into all of its components, and we then will see how each of the components is generated:

The first component on the right hand side of the equation represents the population average *0y* for the trait *y*, and the deviation of each individual’s trait mean from that population average (*Iyi*, generally called individual level or individual intercept). All the other elements occurring on the right hand side of the equation represent also deviations from the average of the trait in that population.

Let’s now go to the end of the equation, where we find the residual *eyhij*. The residual term represents unaccounted effects on the phenotype which could be biological effects and/or measurement errors. At the left of the error term we have *Gyj* that represents the deviation from the population average of trait *y* for a group *j* (this means that all the individuals of a given group *j* share the same value of *G (G* follows a normal distribution with a mean = 0 and variance = *Var(G*)). The concept of grouping is wide ranging from genetic similarities (e.g. families) to spatial aggregations (e.g. territories) and diet compositions and could be specific to the each study.

The last three components (*1y* + S*1yi*).*x1hi* + (*2y* + S*2yi*).*x2hi* + (*12y* + *S12yi*)*x1hi*.*x2hi* represent two different environmental effects and their interaction that you could use to generate different sources of variation for your simulations.

These terms represent phenotypically plastic changes in the value of the trait as a function of two environmental factors *x1* and *x2*. In other words, that part of the equation is the reaction norm of trait *y* as a function of either environment *x1* or *x2* (or both *x1* and *x2*). Below we will explain the equation foragiven environmental factor *x*. Environment *x* can be measured and thus you can decide whether or not to include *x* in subsequent mixed-effects models (whether you include *x* in the model or not allows you to test how unknown/unmeasured environments affect estimates of other parameters in the model). The reaction norm is itself composed of a population-level component (e.g. *1y*) and of an individual deviation from the population norm (*S1yi*). When both terms are combined trait values are generated so that individuals differ in their slopes for trait *y* as a function of environment *x*. Lastly there is a possibility to implement an interaction between *x1* and *x2* [(*12y* + S*12yi*)*x1hi*.*x2hi*)].

**Individual specific responses**

Below we show the way we generate the different values of each individual-level random effect for both traits *y* and *z*. The vector of individual deviations of the population-mean intercept and slopes for the two traits follows a multivariate normal distribution (MNV) with a zero-mean and covariance/variance matrix structure *IS*. On the right hand side, we show the (co)variance matrix *I.* Values of variance stand along the diagonal of the matrix (i.e. respectively, the among-individual variance in intercepts for *y* (*Var(Iy*)), the among-individual variance in slopes for *y* (*Var(Sy)*), the among-individual variance in intercepts for *z* (*Var(Iz)*) and the among-individual variance in slopes for *z* (*Var(Sz)*)). Covariance values between intercepts (e.g., ) slopes (e.g.,), and intercepts and slopes (e.g., ) lie below the diagonal. When building up phenotypic design the program lets you specify all of these (co)variance values. Note that in SQuID inputs the correlation/variance matrix is asked instead of the (co)variance matrix. This is more convenient for users while correlations range between -1 and 1.

**Some of the terms above are explained in more detail in the two following sections (i.e., *Model summary* and *Bivariate model*)**

SQuID can model a bivariate phenotypic equation with random intercepts and slopes for individual identity with respect to two environmental covariates. SQuID enables both linear and interactive effects of these covariates on the bivariate phenotype (i.e., non-warped vs. warped multi-dimensional reaction norm planes; Araya-Ajoy *et al.* 2015; Westneat *et al.* 2015) to be modelled. Covariances in SQuID emerge due to user-defined settings in the fixed and random parts of the equation. Covariances in the random part are set directly by the user, whereas covariances in the fixed part instead emerge when a focal environmental effect is set to affect both traits (e.g., and ). A large number of covariance terms, listed here, can thus be simulated.

|  |  |
| --- | --- |
| Variance component | Explanation |
| Fixed effects |  |
|  | Population-average (i.e., within-individual) covariance between the reaction norm slope for traits and (i.e., linear within-individual response to environmental effect ); exists for all (maximum two) fitted effects (,) |
|  | Population-average (i.e., within-individual) covariance between a reaction norm slope interaction between two environmental effects (,) for trait and the same slope interaction for trait . |
| Random effects |  |
|  | Individual-specific covariance between reaction norm intercepts for traits and |
|  | Individual-specific covariance between a reaction norm slope for trait (i.e., linear within-individual response to environmental effect ) and an intercept for trait ; exists for all (maximum two) fitted effects (,); exists also for the reverse scenario (i.e., ) |
|  | Individual-specific covariance between a reaction norm slope for trait (i.e., linear within-individual response to environmental effect ) and the same slope for trait ; exists for up to two effects (,) (i.e., ,) |
|  | Individual-specific covariance between a reaction norm slope for trait (i.e., linear within-individual response to environmental effect ) and another slope for trait (i.e., response to effect ); exists also for reverse scenario (i.e.,) |
|  | Individual-specific covariance between a reaction norm slope interaction between two environmental effects (,) for trait and an intercept for trait ; exists also for the reverse scenario (i.e., ) |
|  | Individual-specific covariance between a reaction norm slope interaction between two environmental effects (,) for trait and a slope for trait (i.e., linear within-individual response to environmental effect or ), resulting in, respectively, or ); exists also for reverse scenarios (,). |
|  | Individual-specific covariance between a reaction norm slope interaction between two environmental effects (,) for trait and the same slope interaction for trait . |
|  | Higher-order covariance between reaction norm intercepts for traits and |
|  | Residual covariance between reaction norm intercepts for traits and |

**Environments**

SQuID allows you to generate different structures by which the environment changes over time: 1) stochastic

2) temporally auto-correlated

3) linear

4) cyclic.

For the first structure (**stochastic**), environmental values *xSh* are generated using a Normal distribution with mean 0 and variance .

This means that environmental values will be totally stochastic and uncorrelated from one time point to the next. This simple situation is not the most realistic (i.e., it is hard to think of an environmental factor that is completely stochastically distributed in time). The other options are more complex but also more realistic.

For the **temporally** **auto-correlated** *xAh*, we assume that two values of the environment close in time are more similar than two values further apart from each other (i.e., positive autocorrelation). Environmental values *xAh* are generated as the product of a stochastic value by a decay function.

where *h* is the time interval between two instances of the study period, and  is the decay rate, a measure of how fast the correlation decays with time.

where *autocor* is the desired correlation between two successive values.

Once the **linear** trend is set up, SQuID generates a series of values following a classical linear equation.

*xLh* =  *a* + *b*.*h*

where *a* and *b* are respectively, the intercept and the slope of the regression between *xLh* and *h.*  This option allows you to generate environmental values that can change linearly with time, for example mimicking phenology during a season or long-term environmental changes.

If you are interested in mimicking daily or seasonal fluctuations you can choose a **cyclic** (sinusoidal) environmental variation, according to a sinusoidal equation:

*xCh* = *a . sin(b.h + c) +v*,

where|*a*| is the amplitude, 2*Pi*/|*b*| the period, *-c*/|*b*| the horizontal shift, and *v* is the vertical shift. You can therefore change the shape of the cyclic curve by setting the different parameters.

By adding up the different options together you can create environmental data with combinations of effects. Adding the four options together (the most complex situation) will generate environmental data with a linear trend, cyclicity, autocorrelation and stochasticity.

Figure 1: example of 3 environmental effects. (Left) Linear trend with some stochasticity (Middle) Cyclic pattern with linear trend and some stochasticity. (Right) Autocorrelation of 0.7 added to stochastic values.

The environment could be also **shared** or **unshared** among individuals. A shared environment between individuals means that it is general to all the individuals within the population. All individuals will experience the same environment even if the environment would vary over time. In this case, only one environment () will be generated using the different options described above. An unshared environment between individuals means that each individual will experience a specific environment. For instance, if we have 3 individuals in our population 3 different environments will be generated for respectively the individual a, b, and c. Although these environments are different, they follow the same general trend defined by environment options described previously (stochastic, auto-correlated, linear, and cyclic). In order to create multiple environments from one environment configuration, we apply variability around each parameter (e.g. the intercept of the linear effect or the amplitude of the cyclic effect) that follows a normal distribution.

Figure 2: example of 2 environments, one shared (left) and one unshared (right) between 3 individuals. Both environments have been generated with stochastic, linear, and cyclic effect.

Finally, in SQuID the final output of each environmental effect (x1, x2 and x1x2) is expressed in unit variance (i.e. Var(x)=1) and mean-centered (E(x)=0).

**High-level grouping and residual**

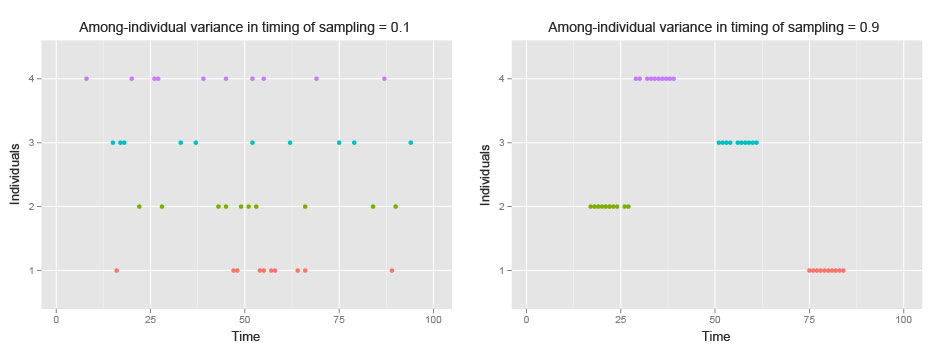
The high-level grouping term (*Gj*) is generated from a normal distribution of mean 0 and variance *VK*, and the error term (*ehij*) from a normal distribution of mean 0 and variance *Ve*.

**SAMPLING DESIGN**

As a second step, we retrieve a subsample of previously simulated phenotypic values according to a specific sampling design. The sampling design between individuals could vary by multiple parameters such as the number of records sampled, and the instance of those records.

First we define the duration of the sampling period that is the same for all individuals.

where is the total simulation duration and is the among-individual variance in timing of sampling (between 0 and 0.95). is used to control the spreading intensity in sampling time between individuals (figure 1). When is small (e.g. 0.1) is large (90% of total simulation duration) and all individuals are sampled almost within the same period. Note that when is 0, all individuals are sampled with the entire period of the simulation. Contrarily, when is large (e.g. 0.9) is small (10% of total simulation duration) and each individual sampling period is spread out within the total simulation duration following a uniform distribution.



**Figure 1**: example of 2 sampling designs where the among-individual variance in timing of sampling is 0.1 (left) and 0.9 (right).

The number of sampled records for each individual (NRi) is generated following a poisson distribution.

where NR is the mean number of records.

Finally, the exact times of sampling for each individual are generated according to a uniform distribution within each individual sampling period.

**INPUT TAB**

In this section you will configure your simulation and sampling design by entering your desired parameter values. All the parameters have default values that you can decide to modify or not.

**Simulation design**

You have to start by generating your world by setting the duration of your study. The default is set to 100 time steps but you can generate less or more time steps. You can also select the number of replicated populations you want to generate (note that data from different populations will be saved in the same output file that you could later use to run statistical analyses). The replicated populations are generated independently using the same simulation design that has been initially inputted by the user. The second step is to decide how many individuals will you create in each sampled population, how many traits you will create for each individual (max = 2), and how many groups of individuals will you define within each population. Note that the number of individuals has to be divisible by the number of groups.

**Environment design**

You have now to decide what types of environmental effects you will generate for the simulated traits. In this section you can select a maximum of two environment effects (*x1* and *x2*). These environments can be customized according to the various options displayed in their respective tabs. You can add stochastic, linear and cyclic effects to the general structured environmental effects. Each option can be added or deleted from the general environmental structure by respectively checking or unchecking its associated checkbox.

For the stochastic effect you have to specify the variance for that environment. This variance will create stochastic variation in the environmental value from one time step to the next. Furthermore, within the stochastic environmental effect section, it is possible to add auto-correlated effects that follow the algorithm presented in the “Description” page. You have to specify the correlation value ranging from 0 to 1 that characterizes the magnitude of temporal autocorrelation.

You can also choose whether or not the environment is showing a linear or a cyclical temporal trend. One more time, you have to specify the parameter values for each option such as the intercept and the slope for the linear effect and the amplitude, the period, the horizontal shift and the vertical shift for the cyclic effect. Therefore, if you select the three options you can create linearly increasing environmental values with some cyclic and stochastic variation from time to time.

For each environment type (stochastic, linear and cyclic), you can also choose whether that environmental effect is shared or not between individuals by checking or not the “Shared environment” checkbox. For instance, you could decide to share the stochastic and linear environmental effects between individuals while the cyclic effect is not shared. When an environment effect is not shared between individuals an input cell appears showing the environmental variance value used to generate the different environments. These individual-specific environments are created by varying their parameters with a normal distribution where the mean is the parameter value itself and the variance is the specified environmental variance and that for each environment effect type.

Furthermore, below each environmental option a graph of the general environmental structure is displayed in order to facilitate the environment visualisation before running the simulation. Note that when the environment is shared (general to all individuals) only one environmental pattern is displayed. In contrast, when the environment is not shared (specific to each individual) multiple environment patterns (one for each individual) are instead displayed, in different colours. To understand how the environment generator works we suggest to start with simple environmental structures such as only linear or cyclic and then add further complexity by combining multiple environmental effect types. Also, remember that in SQuID the final output of each environmental effect (x1, x2 and x1x2) is expressed in unit variance (i.e. Var(x)=1) and mean-centered (i.e. E(x)=0). However, the previsualization graph below displays the generated environmental data before standardization.

Finally, you can add a third environmental effect corresponding to the interaction between *x1* and *x2*. Note that an environmental interaction is allowed only when *x1* and *x2* are selected.

For the analysis part, these environmental effects can be known or unknown to the researcher. For the known environment we assume that you have been able to measure the environmental values and thus will be able to use them for further analyses. Unknown environments, in contrast, represent environments that are not measured. This corresponds to the cumulative effects of the many environmental factors on your trait that you did not have the opportunity to measure but that are still affecting it.

**Model equation**

You are not yet at the stage where you generate the phenotypic values of the trait(s). Note that the program gives you the phenotypic equation of the model you just set up. Any change in the parameters of the model will lead to a change in the equation that is shown below.

**Individual phenotypes design**

The section below allows you to enter all the information on the variance components affecting the phenotypic values of the trait(s). It is a dynamic section that changes according to the parameters you have included earlier in the model. For example when you choose to simulate two traits you obtain a matrix of variance/correlation for the two traits that otherwise does not appear. Depending on the model you have specified earlier some of the matrix cells will appear in colour. These are the variance and correlation components that you can choose to specify according to the model you have chosen before. The section is organised so that you can specify population mean values on top, and then individual variance/correlation values. At the bottom of the section you have the option to add some residual and grouping variance. Please pay attention that in the matrix below, the correlation between each components pair has to be entered and not the covariance.

**Variances summary**

Congratulations! You just got your simulation design configured. At this point you can visualize a summary of all the variances and covariances of the model that you just defined. This summary table contains the value of each variance in addition to its proportion according to the total phenotypic variance.

**Sampling design**

Now you have simulated populations of individuals each with trait values for each time step. It is time to sample as a biologist would by capturing and testing or observing individuals for a limited amount of time during the study period. To mimic this situation SQuID offers several options related to the sampling design you want to use.

In the top left hand side of the sampling design section you have to enter the average number of records per trait. This number has to be between 1 and the number of time steps of the study period. SQuID does not let you specify other numbers. In the input cell underneath you can enter a value for the among-individual variance in timing of sampling (between 0 and 0.95); a variance of zero means that you have no among-individual differences in the timing of sampling. An increase in the variance leads to a decrease in the overlap between individual samples thus a decrease in the time of sampling for each individual.

On the right hand side of the section, there are little check boxes that allow you to indicate whether you want the same or different numbers of records per individual, and if you want individuals to be sampled at the same or different times (when you have two traits you can specify whether you want records for the two traits at the same or different times and the same or different number of records for the two traits). Note that if you click on the box untitled “same sampling time among individuals” SQuID will automatically consider that the number of sampled records is the same between individuals in addition that the records are taken at the same time.

Finally, SQuID offers a previsualization of the sampling design. The figure shows when the trait y of the first 5 individuals will be sampled. The preview figure will be updated automatically when the user changes one of the sampling parameters. Note that the sampling design displayed is just one of many possible examples. We suggest to press the “Refresh” button in order to visualize other possible examples of sampling designs that are generated according to the rules defined by the entered parameters.

**Running the model**

It is now time to run your simulation and to check the output in the output tab. Note that if any inputs have been improperly entered the run button will be inaccessible and it will not possible to run the simulation. For instance, this could happen if you enter a correlation value in the correlation/variance matrix that is out of the authorized range (between -1 and 1) or if the number of sampled records is higher than the duration of the sampling period. In that case you should fix all the input errors and then run the simulation.

**OUTPUT TAB**

While you are running the simulation, SQuID is bringing you to the output tab. Here you have the opportunity to download generated (raw and sampled) data or to inspect figures showing different components of your simulation such as the environment patterns, the individual phenotypic values, and the sampling events over the simulation time.

The first section contains figures that show how the different environmental parameters are distributed in time. When you do not set up any environment parameters the figure show a flat line.

If you click on the “Individual Phenotypes” option, you can access to top figure showing the complete set of values for each individual at each time step, and the bottom one showing the values sampled for each individual according to your sampling design.

Finally by clicking on the “sampling time” option you can see a figure illustrating the distribution of sampling occasions for each individual through time.

We also provide a brief description of the generated data, either raw data or sampled data, which are available to download as a .csv files.