

TABLE OF CONTENTS

	Page
1 Inferring species interactions	1
1.1 Motivation	1
1.2 Methodology	2
1.2.1 Interaction strength	3
1.2.2 Population models	4
1.2.3 Simulation procedure	7
1.2.4 Numerical estimation method	8
1.2.5 Examples	11
1.3 Results	14
1.3.1 Linear model	14
1.3.2 Holling model	17
1.3.3 Range sampling	17
1.4 TEMP : other results	17
1.5 Application to IBM (optional)	18
1.5.1 Testing functional response (and intrinsic growth functions)	20
1.5.2 2 Species IBM model	21
1.5.3 Extend methodology (3, 4 and 5 species)	22
1.5.4 Results	22
1.5.5 2 species	22
1.6 Discussion	22
Bibliography	25

INFERRING SPECIES INTERACTIONS

1.1 Motivation

In this chapter we investigate the possibility of quantifying species interaction strengths from observed population dynamics. This was discussed at length in the introduction (section ??), however we now reiterate some of the important points here and to motivate our approach to the problem.

- metrics for interaction strength (choice of IM metric)
- hare-lynx dataset (ongoing debate!)
- Availability of time-series data. Plankton system (complications! e.g. seasonality)
- Motivate simulation approach (known interactions)
- Timme approach to fit GLV (use of GLV in general - constant interaction strengths)
- Simplify the problem (two species predator prey, but see extension)

We implement and test a novel method for quantifying species interactions from population dynamics. Initially we test this method using two distinct ODE simulation models to generate the population dynamics. These two models are referred to as the *linear* and *holling* models. The GLV can fit the *linear* model exactly. Therefore the results presented in section 1.3.1 serve as a test of our numerical method. We characterise the robustness of the method to the addition of noise, and under sparsity of sampling. The *holling* model cannot be fitted exactly by the GLV because it uses a different form of *functional response*. Therefore applying our method to population dynamics simulated using the *holling* model provides a test of its ability to give approximate

estimates of interaction strength, when underlying structure of the system is not of GLV type. This is also tested in the presence noise and under sparse sampling, and the results are presented in section 1.3.2. In section 1.3.3 we provide a preliminary look at how the method could be used to determine the type of functional response present, by observing the dynamics. Having characterised the performance of our method using ODE models, we then apply it, in section 1.5, to dynamics generated using the IBM model of previous chapters. This represents a first step towards applying the method to empirical data, since it involves a spatial system and a larger number of species. We conclude the chapter with a discussion of how the method could be developed further towards empirical application.

1.2 Methodology

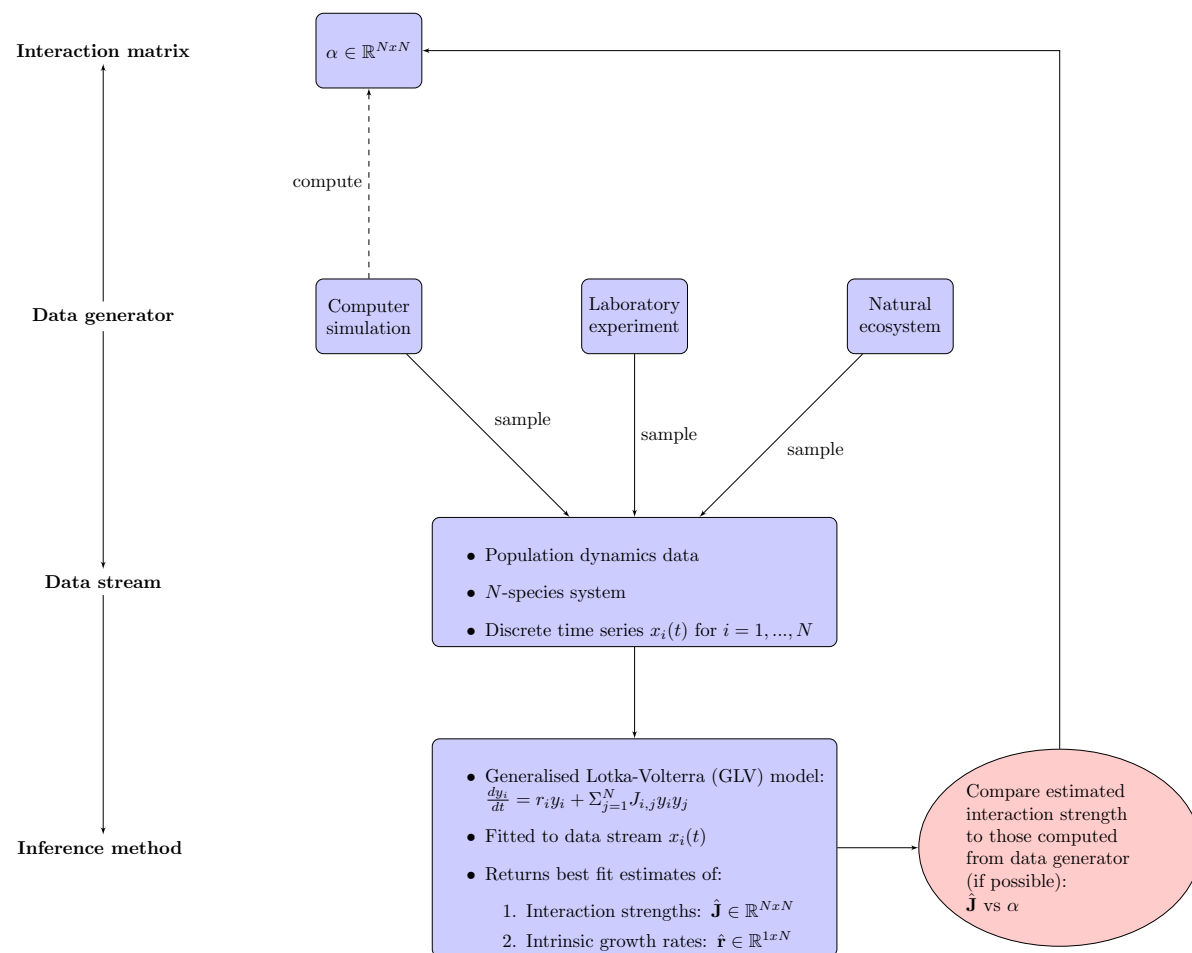


Figure 1.1: This is what we do.

To summarise our methodology, we simulate population dynamics then sample these dynamics and fit a *generalised Lotka-Volterra* (GLV) model. The fitted GLV parameters give us estimates

of the species interaction strengths (and other parameters), which we then compare to those used in the original simulation. The details of all the stages are given below. In section 1.2.1 the *interaction matrix* (IM) is introduced. The IM is the metric used to quantify the strength of species interactions and is key to this chapter. We also introduce the *generalised Lotka-Volterra* (GLV) model, and show that this model has constant interaction strengths, given by the coupling matrix J . In section 1.2.2 we give a general framework for ODE predator-prey modelling, and derive the two models that we use to simulate population dynamics. We then discuss, in section 1.2.3 the details of how these models are simulated *in silico*, including the selection of model parameters. Section 1.2.4 gives the details of the numerical method we use for fitting the GLV model to sampled population dynamics. In section 1.2.5 we give an example of the full methodology in action.

1.2.1 Interaction strength

The metric used for species interaction strength is key to this chapter. As discussed in section 1.1 there are several metrics available [REF]. However there is one that is a natural choice, given our methodology. This metric allows us to calculate the interaction strengths from our simulation models, and to directly compare them to the estimates obtained by fitting the GLV model. The metric is called the *interaction matrix* (IM). The elements of the IM, α_{ij} , quantify the effect of a small change in the population density of species j on the per capita growth rate of species i [REF]. Therefore the IM elements are given by:

$$(1.1) \quad \alpha_{ij} = \frac{\partial}{\partial x_j} \left(\frac{1}{x_i} \frac{dx_i}{dt} \right),$$

where x_i and x_j are the population densities of species i and j respectively. In the case of our two species systems the IM is a 2×2 matrix, but trivially extends to quantify all N^2 pair-wise interactions between species in a N -species system (including self-interactions).

Using the IM we are able to calculate the interaction strengths exactly from the models that we use for simulation. This is because they are ODE models with explicit expressions for dx_i/dt , so we can evaluate the partial derivative in equation 1.1 to obtain analytic forms for all the IM elements ($\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11}$). Depending on the model used the IM elements are either constants, or are functions of prey density. The interaction strengths for our simulation models are given at the end of section 1.2.2, and are illustrated in figure 1.2.

Generalised Lotka-Volterra. The GLV model is the extension of the Lotka-Volterra equations to N species. We use this model to fit to simulated population dynamics and obtain estimates of the underlying interaction strengths. The GLV model is given by:

$$(1.2) \quad \frac{dx_i}{dt} = r_i x_i + \sum_{j=1}^N J_{ij} x_i x_j,$$

where x_i is the population density of species i ; r_i is the intrinsic growth rate; N is the number of species and J_{ij} is the coupling between species i and j . Applying equation 1.1 to equation 1.2 we see that $\alpha_{ij} = J_{ij}$, such that the IM for the GLV model is equal to the coupling matrix. Therefore, by fitting the GLV model to population dynamics, the fitted parameters J_{ij} give us numeric estimates of interaction strength. To perform the fit we use the method detailed in section 1.2.4.

1.2.2 Population models

We use ordinary differential equation (ODE) models to simulate two species predator-prey dynamics. Therefore the mathematics presented in this section focuses on this case. However the framework may be easily extended to include other interaction types (competition and mutualism), and to model larger systems in a similar way. For a two species system the dynamics are governed by two coupled equations, one for each species. Importantly it is possible, using the IM introduced in section 1.2.1, to calculate the species interaction strengths from the model. The model equations take the general form:

$$(1.3) \quad \begin{aligned} \frac{dx_0}{dt} &= G_0(x_0) + a_{01}x_1H(x_0, x_1), \\ \frac{dx_1}{dt} &= G_1(x_1) + a_{10}x_1H(x_0, x_1) \end{aligned}$$

where species x_0 and x_1 are the population densities of the prey and the predator species respectively; the $G_i(x_i)$ are the intrinsic growth functions of each species; the a_{ij} are constant coefficients and $H(x_0, x_1)$ is the functional response (FR) of the predator. This form is standard in the literature [REFS] and many models may be expressed in this way by choosing different functional forms for G and H . The coefficients a_{01} and a_{10} are negative and positive respectively, such that the prey losses biomass, and the predator gains biomass as a result of the interaction. These coefficients may be used to introduce asymmetry into the interaction terms. For example it is common to choose $|a_{01}| > |a_{10}|$, to model the inefficiency of the predator in the conversion of biomass from the prey. For the intrinsic growth functions we use the functional forms:

$$(1.4) \quad G_0(x_0) = r_0x_0 \left(1 - \frac{x_0}{K_c}\right)$$

$$(1.5) \quad G_1(x_1) = -r_1x_1,$$

where $r_i \in \mathbb{R}^+$ is the intrinsic growth rate of species i , and K_c is the carrying capacity of the prey species. Therefore the predator has an exponential intrinsic mortality, whereas the prey species has logistic intrinsic growth. These use of these functional forms in predator-prey modelling was made popular by Rosenzweig and MacArthur [REF]. They are now widely used [REFS].

The FR defines the per-predator rate of consumption of prey. We focus on the forms proposed by Holling in the 1950s [REFs], which are also widely used [REFS]. However it is worth noting that various other forms have been proposed and there is an ongoing debate about which to use [REFS] (see discussion in section 1.6). There are three types of Holling FR, referred to as types I, II, and III. These can be expressed as:

$$(1.6) \quad H_I(x_0, x_1) = x_0,$$

$$(1.7) \quad H_{II}(x_0, x_1) = \frac{x_0}{x_0 + K_s},$$

$$(1.8) \quad H_{III}(x_0, x_1) = \frac{x_0^2}{x_0^2 + K_s^2},$$

where x_0 is the prey density, and K_s is the saturation constant for the predator, giving the prey density at which the per-predator consumption rate reaches half-maximum. We choose to narrow our investigation by focusing here on the first two forms: Holling type I and type II. Therefore we have two distinct simulation models, which we refer to as the *linear* and *holling* models.

Linear model. This model uses the type I FR. As is clear from equation 1.6, type I is the simplest of the Holling functions: the per-predator predation rate increases linearly as the abundance of available prey increases. The slope of this linear relationship given by the a_{ij} parameters in equations 1.5. The linear FR is the same as is used in the famous Lotka-Volterra equations [REF], meaning that our linear model may be expressed in GLV form (equation 1.2). We may rescale the equations of the linear model in order to reduce the number of parameters. This makes the local stability analysis simpler, and reduces the dimension of the search space when probing the equations numerically via simulation. The re-scaled equations are given by:

$$(1.9) \quad \frac{d\chi_0}{dt} = A\chi_0(1 - \chi_0) - B\chi_0\chi_1$$

$$(1.10) \quad \frac{d\chi_1}{dt} = -\chi_1 + C\chi_0\chi_1,$$

where χ_0 and χ_1 are the re-scaled prey and predator population densities respectively; the parameters $A, B, C \in \mathbb{R}^+$. The equilibrium population densities are given by:

$$(1.11) \quad \chi_0^* = \frac{1}{C}$$

$$(1.12) \quad \chi_1^* = \frac{A}{B} \left(1 - \frac{1}{C}\right),$$

Therefore χ_0^* is always positive, and χ_1^* is positive if $c > 0$. This is a requirement for physical realism, since it is not possible to have negative populations of species. In most applications it

is also required that this equilibrium is stable, to allow for the coexistence of species. We use these conditions on the equilibrium for parameter selection, which is discussed in section 1.2.3. By applying equation 1.1 to equations 1.9 and 1.10 we can evaluate the elements of the IM for the linear model. This gives:

$$(1.13) \quad \alpha_{linear} = \begin{bmatrix} -A & -B \\ C & 0 \end{bmatrix},$$

such that all the interaction strengths are constants, which is illustrated in figure 1.3.

Holling model. This model uses the type II FR, which is a non-linear function of prey density. As we can see from equation 1.7 and figure 1.2, this FR models predator saturation - individuals take a certain amount of time to process and digest prey - such that the response curve flattens out at high prey densities. The difference between the type I and II functions is illustrated in figure 1.2. We may perform a similar rescaling as we did with the linear model to reduce the number of parameters. The resulting equations for the holling model are given by:

$$(1.14) \quad \frac{d\chi_0}{dt} = A\chi_0(1-\chi_0) - \frac{B\chi_0\chi_1}{\chi_0 + D}$$

$$(1.15) \quad \frac{d\chi_1}{dt} = -\chi_1 + \frac{C\chi_0\chi_1}{\chi_0 + D},$$

where the saturation constant $D \in \mathbb{R}^+$, and the other symbols are the same as in equations 1.9 and 1.10. The equilibrium populations for this model are given by:

$$(1.16) \quad \chi_0^* = \frac{D}{C-1}$$

$$(1.17) \quad \chi_1^* = \frac{ACD(C-1-D)}{B(C-1)^2},$$

Therefore χ_0^* is positive if $C > 1$, χ_1^* is positive if $C - D > 1$. Again we use these conditions, and the requirement of stability, to constrain our choice of parameters (1.2.3). As we did for the linear model, we can evaluate the IM for the holling model, giving:

$$(1.18) \quad \alpha_{holling} = \begin{bmatrix} -A + \frac{B\chi_1}{(\chi_0+D)^2} & \frac{-B}{\chi_0+D} \\ \frac{CD}{(\chi_0+D)^2} & 0 \end{bmatrix},$$

such that the three non-zero elements of the IM are functions of prey density χ_0 , instead of constants. The shape of these interaction functions is shown in figure 1.2, and we will return to them in section 1.2.5.

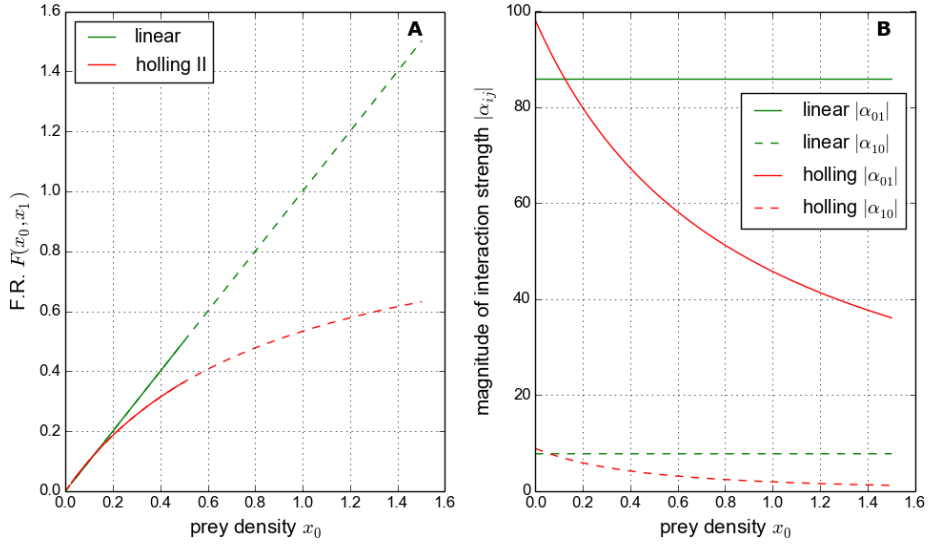


Figure 1.2: Example of (A) the functional response curve, and (B) the corresponding inter-specific interaction strengths for one parameter set of the *linear* model, and one of the *holling* model.

	A	B	C	D
linear	0.1 - 100	0.1 - 100	1 - 100	N/A
holling	0.1 - 100	0.1 - 100	1.1 - 100	0.1-99

Table 1.1: Ranges from which parameters were selected uniformly at random for the two ODE simulation models. The parameters are all allowed to vary over at least three orders of magnitude, to ensure that our investigation covers a large region of parameter space. The restrictions on parameters C and D ensure that it is always possible to achieve an equilibrium population of both species that is strictly positive (see equations 1.11, 1.12, 1.16, 1.17)

1.2.3 Simulation procedure

We apply a strict recipe when running simulations in order to ensure consistency and to allow comparison of our numerical results. Key to this is the control of certain variables across simulations, and also our method for parameter selection, both of which are discussed below. All simulations are run using the first-order forward Euler approximation to the ODE model. We use additive Gaussian noise to simulate *process error*. Therefore we implement the stochastic difference equation :

$$(1.19) \quad \chi_{i,t+1} = \chi_{i,t} + \Delta t \Delta \chi_{i,t} + \xi_{i,t},$$

where Δt is the integration time step, $\Delta \chi_i$ is given by the right hand side of ODE model being simulated (e.g. equations 1.9, 1.10 for the linear model); and the additive noise term

$\xi_{i,t} \sim \mathbb{N}(0, \sigma_{noise} \chi_{i,t} \Delta t)$. The value of σ_{noise} is quoted as *noise intensity* in what follows. In the event of stochastic extinction of either species, both population densities are reset to their initial conditions. The case where $\sigma_{noise} = 0$ is referred to as the *deterministic* model. In all the results presented the simulations were run with a time step $\Delta t = 10^{-4}$. All code was implemented in the language *Python*, and large computations were performed on the UoB HPC cluster *Blue Crystal* [REF].

The goal of fitting the GLV model to simulated population dynamics requires that the dynamics contain enough information to perform the fit - it is not possible to fit the a model if species populations are sitting at equilibrium. Therefore we follow the precedent set in [REF], such that all simulated dynamics of the *deterministic models* exhibit two ‘large amplitude’ oscillations about a stable equilibrium (see condition 2 below). Every simulation is run with the initial population densities set to half of their equilibrium value. This ensures that all systems start consistently away from equilibrium.

Parameter selection. We select an ensemble of 100 parameter sets for both simulation models (*linear* and *holling*). Parameters are selected uniformly at random from predefined ranges, which are given in table 1.1. This range ensures that a positive equilibrium population is possible (see equations 1.11, 1.12, 1.16, 1.17), but also allows for parameters to vary over at least three orders of magnitude so that our investigation covers a large region of parameter space. The selected parameters are then accepted if they meet the following conditions:

1. The equilibrium population is positive, and is locally a stable spiral (eigenvalues of the Jacobian have negative real part and complex conjugate imaginary part).
2. The deterministic dynamics exhibit at least two full rotations in the phase plane before relaxing to within 5% of the equilibrium (Euclidean distance in the phase plane).
3. The population densities do not differ by more than an order of magnitude, in the deterministic case.

The two parameter sets generated by the above procedure are used for the simulation results presented in section 1.3. All simulations, including those with $\sigma_{noise} \neq 0$, are run for the length of time T_{2P} required to achieve two full oscillations in the deterministic case, for that parameter set.

1.2.4 Numerical estimation method

To estimate the inter-specific interaction strengths we use a numerical method, adapted from [1], to fit the GLV model to the population dynamics. The method gives ‘best fit’ estimates of the GLV parameters, which include constant coefficients for the interaction strengths as we saw in section 1.2.1. We include here a derivation of the method, slightly adapted and simplified for

our purposes. Say the dynamics of the population density of a species x_i is governed by coupled differential equations of the form:

$$(1.20) \quad \dot{x}_i = r_i f_i(x_i) + \sum_{j=1}^N J_{i,j} g_{ij}(x_i, x_j),$$

where $\dot{x}_i = \frac{d}{dt}x_i$; N is the number of species in the system and i, j index the species. The $r_i, J_{ij} \in \mathbb{R}^+$ are constants, and the functions f_i and g_{ij} are known. This form looks familiar, indeed all of the ODE models discussed so far may be expressed in this form - there is an intrinsic growth term, and a linear sum of interaction terms. To express the GLV model (equation 1.2) in this form, we have:

$$(1.21) \quad f_i(x_i) = x_i$$

$$(1.22) \quad g_{ij}(x_i, x_j) = x_i x_j,$$

It would be possible to use this method to fit models other than the GLV, so long as the functions f_i and g_{ij} are *known and parametrised*. Since the functions f_i and g_{ij} are known there are $N + 1$ unknowns in equation 1.20: r_i and $J_{i,j}$ for $j = 1, \dots, N$. Therefore, if we knew the exact values of \dot{x}_i, x_i and the x_j 's, at $N + 1$ time points, then we could solve the equation for r_i and the $J_{i,j}$'s. However in any practical application our knowledge of the system is not *exact*; the system is subject to noise; and the model may be an imperfect description of the dynamics. So the equation cannot be solved exactly. We must look for an approximate solution. To do this the full state of the system is sampled at $M + 1$ time points t_m for $m \in 1, \dots, M, M + 1$. These samples are used to construct estimates for the states x_i and their time-derivatives \dot{x}_i at M intermediate time points, for every species i .

The simplest way to estimate the time-derivatives is to take the finite difference between observations at two consecutive time points, giving estimates:

$$(1.23) \quad \hat{\dot{x}}_i(\tau_m) := \frac{x_i(t_m) - x_i(t_{m-1})}{t_m - t_{m-1}},$$

where $\tau_m \in \mathbb{R}, m \in \{1, \dots, M\}$ is the midpoint of the two time-points:

$$(1.24) \quad \tau_m := \frac{t_{m-1} + t_m}{2}.$$

To evaluate the functions f_i, g_{ij} at these new time-points we must estimate the states $x_i(\tau_m)$ from our observations. We use the linear interpolation:

$$(1.25) \quad \hat{x}_i(\tau_m) := \frac{x_i(t_{m-1}) + x_i(t_m)}{2}.$$

So from equation 1.20 we can now construct M equations using our $M + 1$ samples:

$$(1.26) \quad \hat{x}_i(\tau_m) = r_i f_i(\hat{x}_i(\tau_m)) + \sum_{j=1}^N J_{i,j} g_{ij}(\hat{x}_i(\tau_m), \hat{x}_j(\tau_m)).$$

We now simplify the notation such that equation 1.26 may be written

$$(1.27) \quad \hat{x}_{i,m} = r_i f_{i,m} + \sum_{j=1}^N J_{i,j} g_{ij,m},$$

where the subscripts i, j indicate the species, and m indicates the time-point τ_m for which the equation holds. This system of M equations can be expressed in matrix form:

$$(1.28) \quad X_i = J_i G_i,$$

where we have

$$(1.29) \quad X_i = \begin{pmatrix} \hat{x}_{i,1} & \hat{x}_{i,2} & \cdots & \hat{x}_{i,M} \end{pmatrix} \in \mathbb{R}^{1 \times M},$$

$$(1.30) \quad J_i = \begin{pmatrix} r_i & J_{i,1} & J_{i,2} & \cdots & J_{i,N} \end{pmatrix} \in \mathbb{R}^{1 \times (N+1)},$$

$$(1.31) \quad G_i = \begin{pmatrix} f_{i,1} & f_{i,2} & \cdots & f_{i,M} \\ g_{i,1,1} & g_{i,1,2} & \cdots & g_{i,1,M} \\ g_{i,2,1} & g_{i,2,2} & \cdots & g_{i,2,M} \\ \vdots & \vdots & \ddots & \vdots \\ g_{i,N,1} & g_{i,N,2} & \cdots & g_{i,N,M} \end{pmatrix} \in \mathbb{R}^{(N+1) \times M}.$$

The system 1.28 has $N + 1$ unknowns ($J_{i,k}$ for $k = 1, \dots, N + 1$) and M equations. In the case when $M > N + 1$ the system is over constrained and there is no exact solution in general. We look for an approximate solution \hat{J}_i that minimises the error between the LHS and RHS of equation 1.28. We take the error function:

$$(1.32) \quad E_i(\hat{J}_i) = \sum_{m=1}^M (X_{i,m} - \sum_{k=1}^{N+1} \hat{J}_{i,k} G_{i,k,m})^2,$$

which we want to minimise with respect to the matrix elements $\hat{J}_{i,k}$. That is

$$(1.33) \quad \frac{\partial}{\partial \hat{J}_{i,k}} E_i(\hat{J}_i) \stackrel{!}{=} 0.$$

By taking the derivative of the RHS of equation 1.32 we have that:

$$\begin{aligned}\frac{\partial}{\partial \hat{J}_{i,k'}} E_i(\hat{J}_i) &= \frac{\partial}{\partial \hat{J}_{i,k'}} [\sum_{m=1}^M (X_{i,m} - \sum_{k=1}^N \hat{J}_{i,k} G_{i,k,m})^2] \\ &= -2 \sum_{m=1}^M [(X_{i,m} - \sum_{k=1}^N \hat{J}_{i,k} G_{i,k,m}) G_{i,k',m}]\end{aligned}$$

To find the minimum of the error function we equate this to zero, giving:

$$\begin{aligned}0 &= \sum_{m=1}^M (-X_{i,m} G_{i,k',m} + G_{i,k',m} \sum_{k=1}^N \hat{J}_{i,k} G_{i,k,m}) \\ &= (-X_i G_i^T)_{k'} + \sum_{m=1}^M G_{i,k',m} (\hat{J}_i G_i)_m \\ &= (-X_i G_i^T)_{k'} + \sum_{m=1}^M (\hat{J}_i G_i)_m G_{i,m,k'}^T \\ (1.34) \quad &= -X_i G_i^T + \hat{J}_i G_i G_i^T\end{aligned}$$

Therefore we conclude that:

$$(1.35) \quad \hat{J}_i = X G_i^T (G_i G_i^T)^{-1},$$

which, in our case, is the analytic form for the best estimate of the row corresponding to species i in parameter matrix of the GLV model. For a two species system, by applying equation 1.35 to each species, we can obtain the full set of GLV parameter estimates:

$$(1.36) \quad \hat{J} = \begin{pmatrix} \hat{J}_{0,0} & \hat{J}_{0,1} \\ \hat{J}_{1,0} & \hat{J}_{1,1} \end{pmatrix},$$

and

$$(1.37) \quad \hat{r} = \begin{pmatrix} \hat{r}_0 & \hat{r}_1 \end{pmatrix}.$$

We perform the computation by constructing the matrices X_i and G_i for both species, and do the matrix multiplication using the *Python* package *numpy*[REF]. The fact that the error minimisation has an analytic solution makes it very computationally efficient, allowing us to perform many replicate calculations. However the performance of the method may be lower than other, more computationally expensive, model fitting algorithms (see discussion section 1.6). It is possible to assess the goodness of fit achieved by evaluating the error function (equation 1.32).

1.2.5 Examples

Here we show examples of the dynamics of both models, both with and without noise. For the holling model we also show the variability in interaction strengths during the simulations. We also present a table with the results of the GLV, comparing them to the simulation interaction strengths.

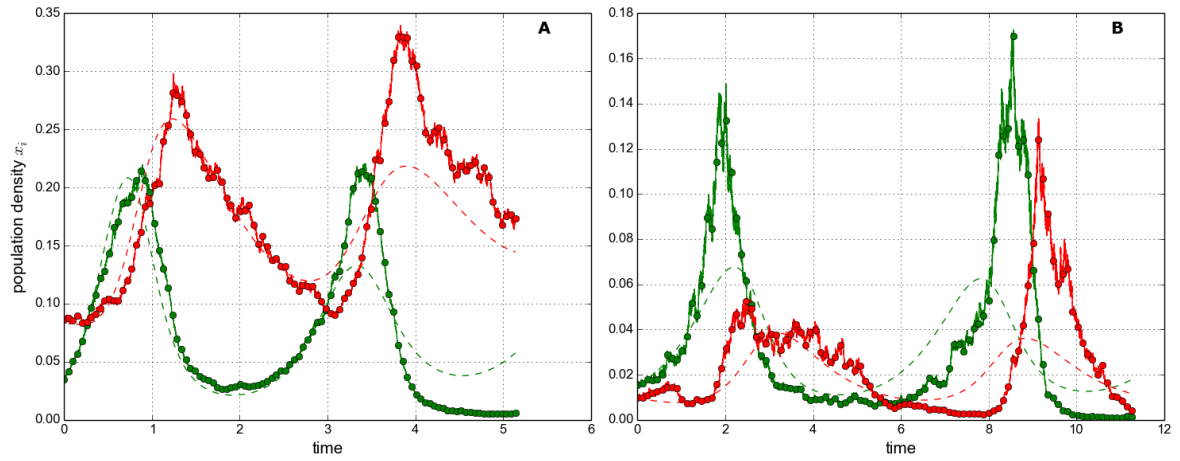


Figure 1.3: Example linear dynamics. 100 sampels. Two different parameter sets. A: noise=20. B:noise=50.

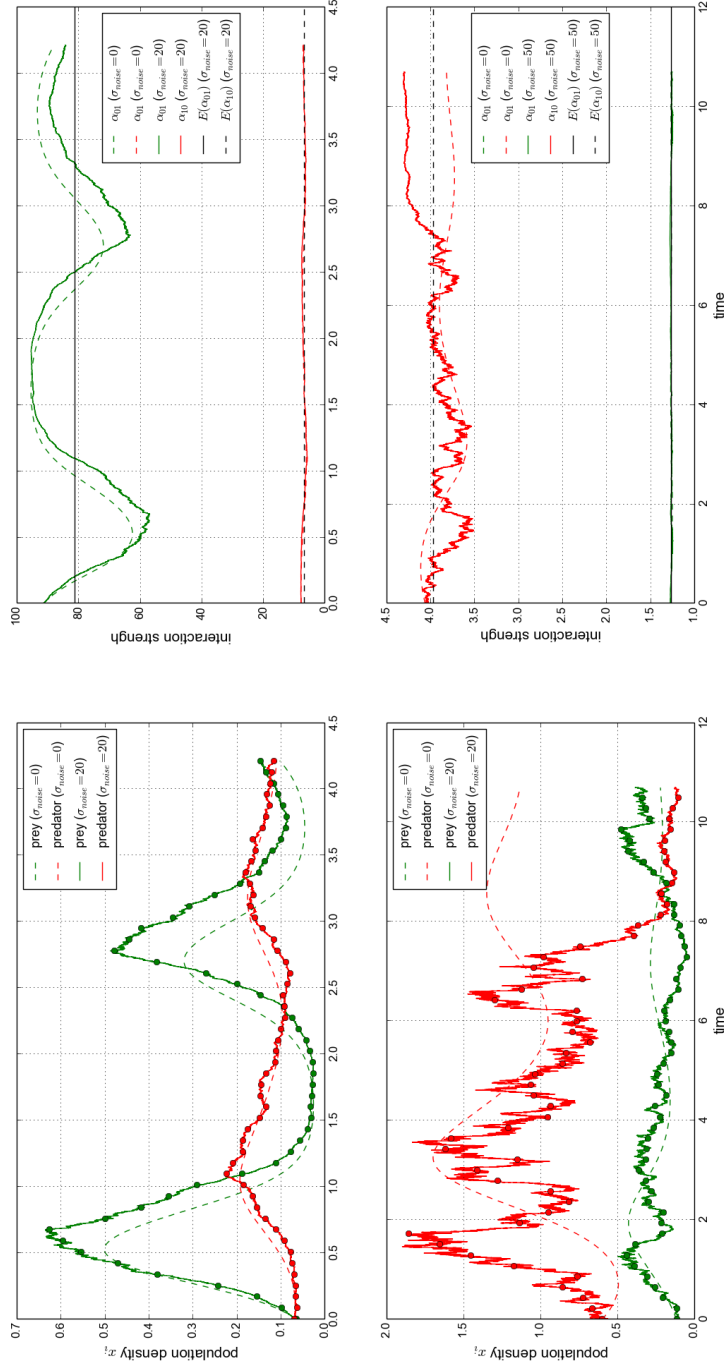


Figure 1.4: Example HOLLING II dynamics.

1.3 Results

In this sections we characterise the numerical performance of the method, described in section 1.2.4, for estimating the strength of interactions between species. The method is tested on the dynamics of two different ODE systems: a Lotka-Volterra (LV) and a Holling type II (HII) system. In the first case it is simply a test of a model fitting procedure. This is because the method works by fitting a generalised Lotka-Volterra (GLV) model to the dynamics, and the LV systems can be expressed as a GLV systems. Therefore we are simply simulating using one model, and then testing a method of estimating the model parameters from the simulated dynamics. We test the effects of noise and sampling frequency. In the second case, the HII system cannot be expressed as a GLV system. Therefore the GLV model that we fit can only approximate the dynamics and we cannot make a direct comparison between the parameters of the simulation model and the GLV model used for estimation. In this case we compare to the mean interaction strengths (see section 1.3.2).

1.3.1 Linear model

Initially we run repeated simulations of the LV model using a single parameter set. We investigate how the numerical estimates of the model parameters respond to two variables: the level of noise in the simulations; and the number of samples used for estimation. Other variables are held constant using the simulation procedure described in section 1.2.3. We then generalise these results by looking at the relative error in the estimates, for repeated simulations using an ensemble of 100 selected parameter sets (as described in section 1.2.3).

Single parameter set. Here we can make direct comparison between model parameters. The GLV model for two species has six parameters: $r_0, r_1, J_{00}, J_{01}, J_{10}, J_{11}$. These correspond respectively to the following constant values of the LV system used for simulations: $A, -1, -A, -B, C, 0$ (see equation ??). In general we find that the numerical estimates perform well at low noise intensities and poorly at high noise intensities. This is illustrated in figures 1.5 and 1.6. We also find that the estimates improve with the number of samples used, up to a point. Beyond this point the use of more samples does little to improve to estimates, and in some cases makes them worse. This behaviour is illustrated in figures 1.7 and 1.8. These patterns were found to hold across all parameter sets investigated, but are only shown using a single parameter set here for clarity.

In panel **A** of figures 1.5 and 1.6 we see that the mean value of the estimates approaches the true value for low noise and, in panel **B** that the variance in the estimates approaches zero. This tells us that the method consistently gives a good fit of the GLV to the dynamics of the LV system, even when only 100 sample points are used (figure 1.5). As the noise intensity is increased the mean values of the estimates deviate from the true values, and the standard deviation in the estimates increases. Comparing the two figures we see that the response to noise is very similar

whether 100 or 10,000 samples are used. A notable exception to this is a spike in the variance in panel **B** of figure. However this appears to be a single statistically anomolous result and not part of the trend. Panel **C** of both figures shows that the error function, which is minimised by the estimation method, increases with noise for both species. This cannot be directly compared between the two plots because of the different number of samples used. However it indicates that in both cases (100 and 10,000 samples) the quality of the fit is high in the deterministic case, and decreases with noise.

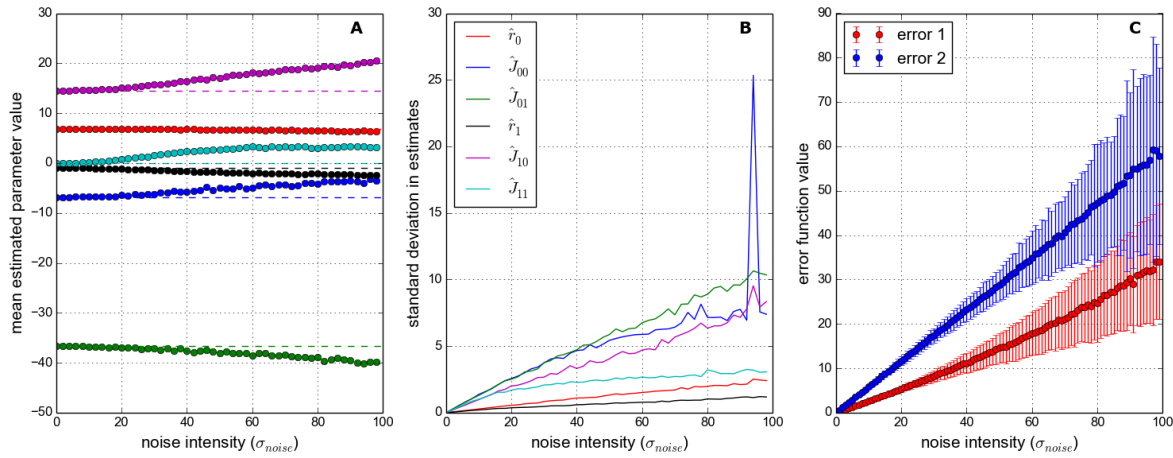


Figure 1.5: Effect of noise on numerical estimates. Here the method uses 100 samples from simulated dynamics. All simulations run using the LV model with a single parameter set. The noise intensity varies between 0 (deterministic) and 100. See section 1.2.5 for an intuition of how noisy this is. 1000 repeat simulations run at each noise level. **Panel A:** Mean estimated parameter values (each dot representing mean over 1000 repeats). The ‘true’ paramter values of the simulation model are shown by dashed lines. **Panel B:** Standard deviation in estimates. **Panel C:** Value of the error functions used in the estimation method, one for each species. The dots show the mean error, and the bars show \pm one standard deviation.

We now look at how the estimates respond to the number of samples used, in the cases of low and high noise intensity. Figure 1.7 shows the low noise case, with $\sigma_{noise} = 10$. In panel **A** we see that the mean value of the estimates quickly converges to close to the true parameter values, as the number of samples increases. Panel **B** shows that the standard deviation in the estimates quickly becomes small, but non-zero. Above about 32 samples there is no visible improvement in the estimates, as measured by the mean or standard deviation. In figure 1.8 we see the effect of a higher noise intensity. Here we have $\sigma_{noise} = 50$. Panel **A** shows that the estimates do not converge on the true parameter values, even for large numbers of samples. Also the standard deviation in the estimates, shown in panel **B**, is higher than in the low noise case. Again we find that there is little, if any, improvement in the estimates beyond about 32 samples.

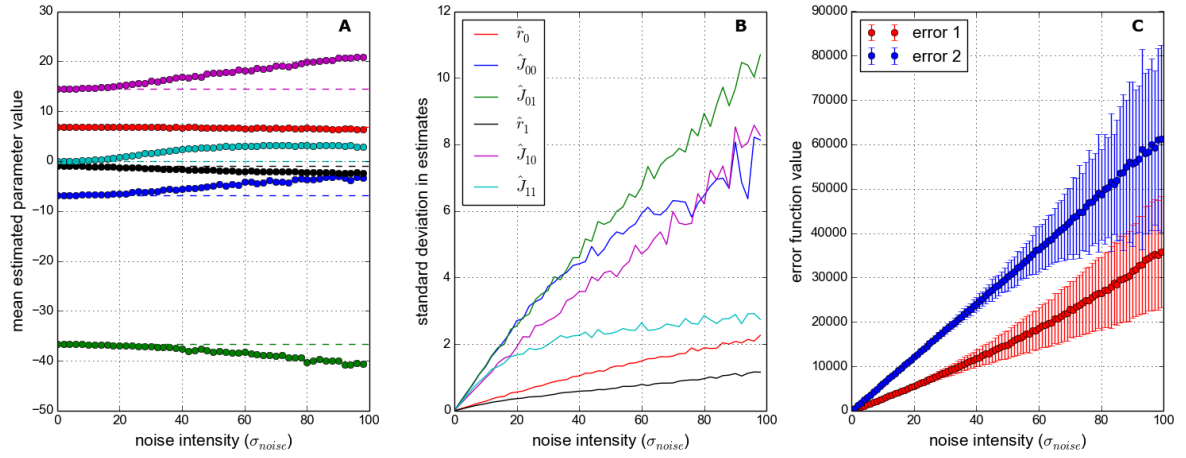


Figure 1.6: Exactly as in figure 1.5 but using 10,000 samples from the simulated dynamics.

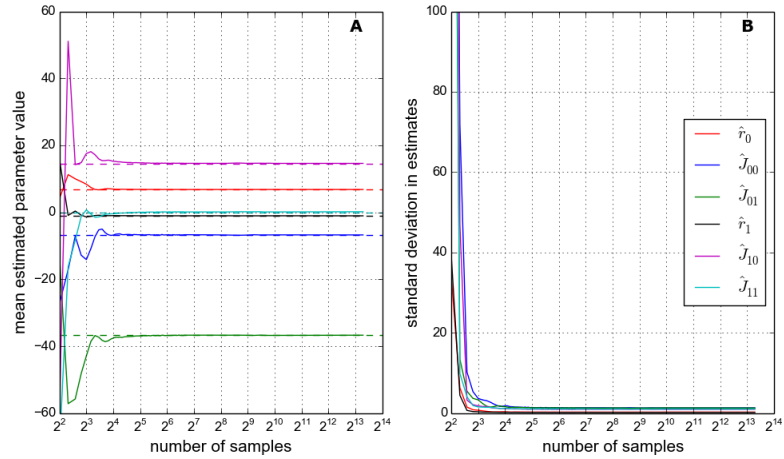


Figure 1.7: Effect of the number of samples on numerical estimates. All simulations run using the LV model with a single parameter set. The noise intensity $\sigma_{noise} = 10$. Number of samples ranges from 4 to 10,000. Samples drawn from simulated dynamics at equal intervals. 1000 repeat simulations for each number of samples. **Panel A:** Solid lines show mean estimated parameter values. Dashed lines show the ‘true’ parameter values of the simulation model. **Panel B:** Standard deviation in estimates.

Ensemble of parameter sets. Run 10 repeats for each of 100 parameter sets. In general the trends described above hold across the ensemble..

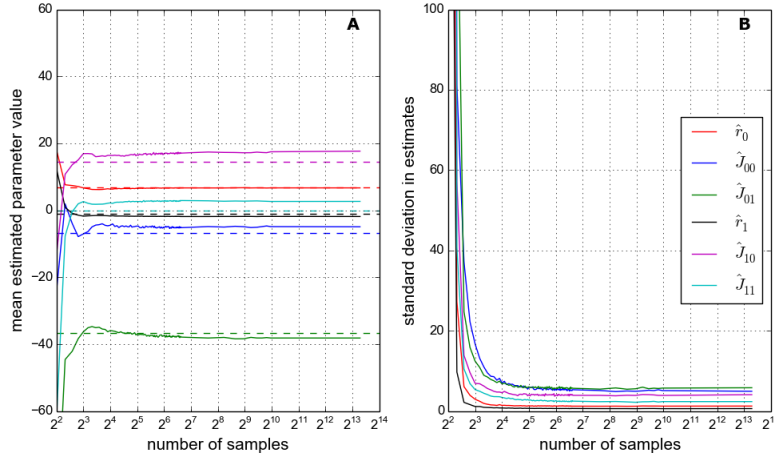


Figure 1.8: Exactly as in figure 1.7 but with noise intensity $\sigma_{noise} = 50$.

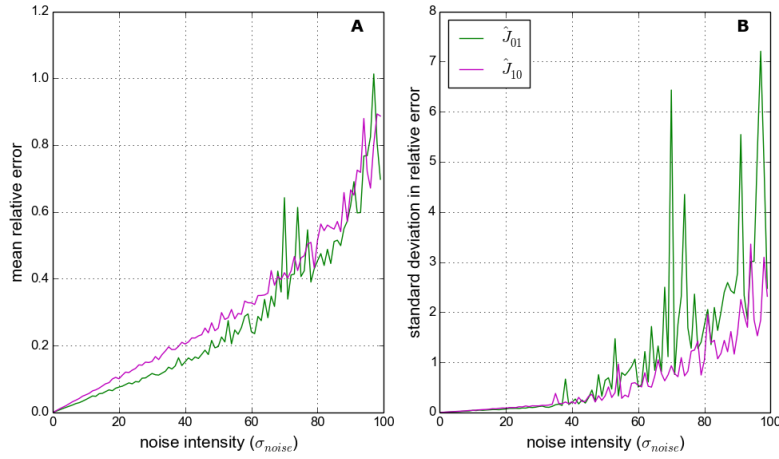


Figure 1.9: Nonsense. 1000 samples used.

1.3.2 Holling model

1.3.3 Range sampling

1.4 TEMP : other results

This section shows some plots which I was not planning to put into the thesis but are worth discussing..

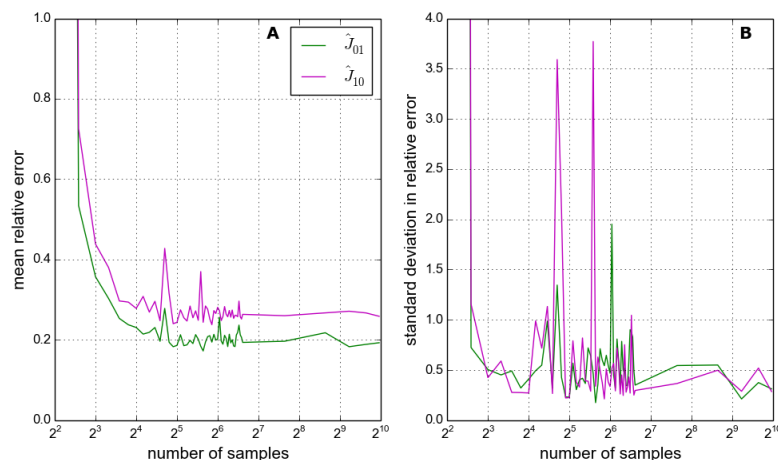


Figure 1.10: Nonsense. Noise is 50.

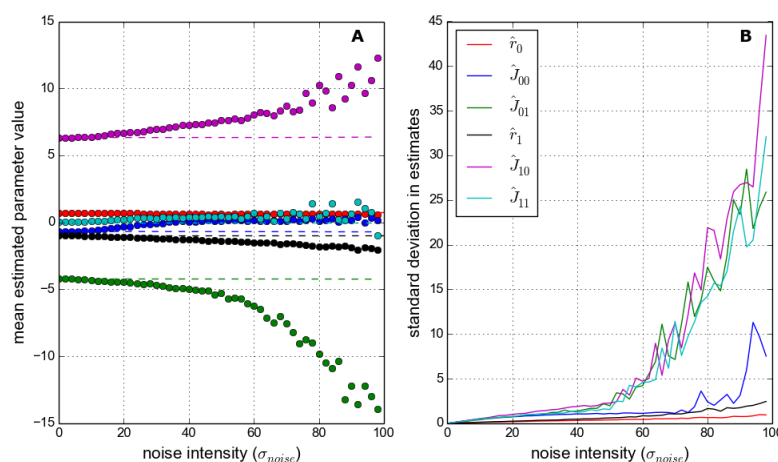


Figure 1.11: Nsamples is 10,000.

1.5 Application to IBM (optional)

In this section we apply the methodology for inferring species interactions to the IBM simulations. In the previous section we have seen that the method works well when fitting the GLV to two species predator-prey dynamics simulated with ODE models. In the case that the dynamics is governed by the Lotkva-Volterra equations, and in the absence of noise, fitting the GLV model produces true estimates of the underlying parameters which include the inter-specific interaction strengths. The estimates require relatively few samples in order to achieve high accuracy, and which converge on the true parameter values as the sampling intensity increases. However as noise is added to the simulations the accuracy of the estimates decreases. In particular we found that, in the presence of noise, the estimates do not converge on the true parameter values -

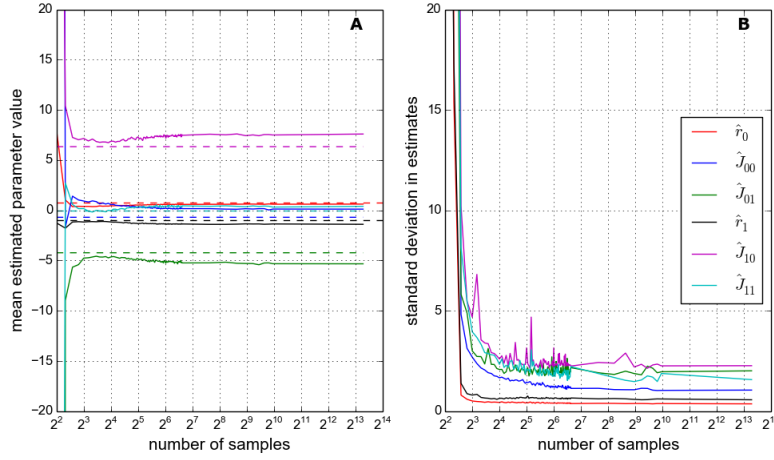


Figure 1.12: Noise is 50.

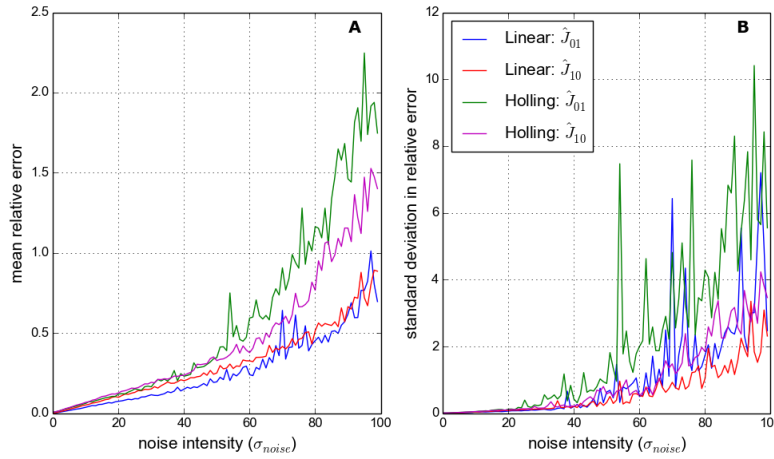


Figure 1.13: Noise is 50.

that is, noise introduces systematic error in to the estimates. In the case that the dynamics are governed by the Holling II model matters are complicated - but in general we find that the GLV can approximately capture the strength of species interactions (and also the dynamics?), provided there is not too much noise, and the FR is not too non-linear¹.

Can the IBM dynamics be approximated by the GLV model? The hypothesis is that it can. Argue this..and refer to previous mention of LV type dynamics. Refer forward to testing linearity of FR. Exponential growth and decay, linear FR. However there are certain problem/factors that may hinder this approach/represent a departure from GLV dynamics...Noise, spatial effects, bioenergetic model - time delay? Immigration (one component of noise).

The issue of noise is important since, as we have shown in chapter REF, there is a strong

¹This is all need to be demonstrated - WORK TO DO!

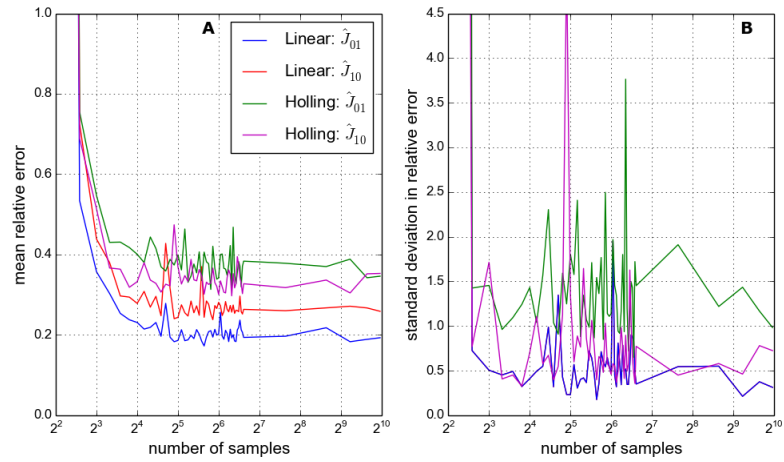


Figure 1.14: Noise is 50.

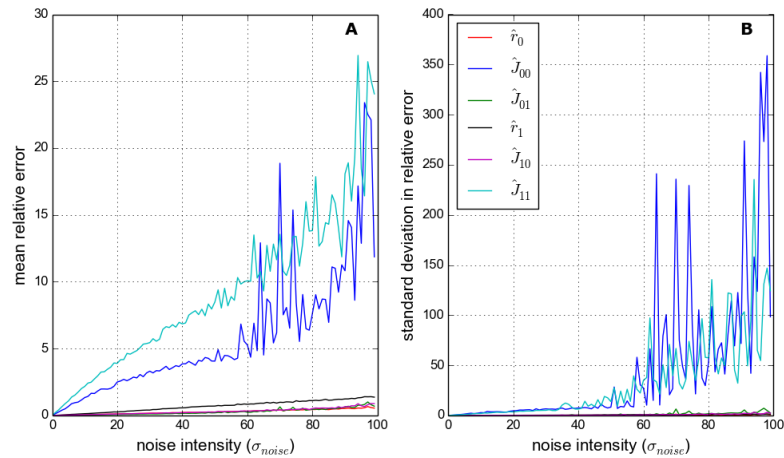


Figure 1.15: Nonsense. 1000 samples used.

stochastic component to the IBM simulations.

Modelling individuals, not biomass or energy. Is this problematic? Set herb-fraction to 1. Other issues?

1.5.1 Testing functional response (and intrinsic growth functions)

Here we demonstrate the linearity of the predator functional response for animal predators in the second and third trophic levels (using 2sp and 3sp chain) - demonstrate that there is a slight difference. But basically linear. Other issues that may arise - high noise and low abundances create error.

We also conduct an experiment to test the intrinsic growth and death rates of plant and animal species - by setting herb-fraction=0.0. Demonstrate they are well approximated by exponential.

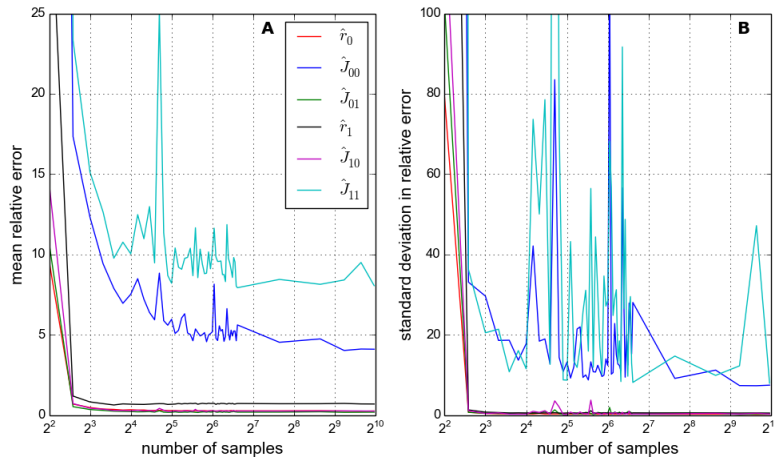


Figure 1.16: Nonsense. Noise is 50.

In the absence of immigration.

Carrying capacity: is there evidence for density dependent birth/death? - use this fact in later analysis. Also discuss that the carrying capacity will vary with the number of species, not just a single species thing (non-pairwise interactions in competition for space - argghh!)

1.5.2 2 Species IBM model

Define the model and what the inferred parameters represent:

- J_{01} : per-capita of rate consumption of the prey
- J_{10} : per-capita of rate reproduction of predator, due to consumption of prey. Not as well defined. But only source of predator births? Numerical response! (get REF)
- J_{00} : intra-specific regulation of prey growth - see carrying capacity experiment.
- J_{11} : intra-specific regulation of predator mortality? Check this. Expect zero? Or expect high number of predator means more reproduction because easier to find partner, therefore reduce mortality? Or increase birth rate. Not clear. Again SEE EXPERIMENT.
- r_0 : prey intrinsic growth rate. Estimate from exp?
- r_1 : predator intrinsic mortality rate. Estimate from exp?

So we know what values to expect, or at least the signs. We can evaluate the model fit by comparing the values of these estimates with birth/death rates from the simulations. Although not totally fair. We can also simulate GLV with the inferred parameters - does it match. Is the equilibrium the same? And check the error function of the fit. We show all this in the results section below.

Dynamics of the model we two species is a new thing. Here we show that with the default parameters we get relaxation-type oscillations. This is interesting, we try fitting to these in the two species case. But they become problematic for larger number of species. Argue why. Therefore we increase the reproduction rate (refer to previous chapter), which creates more WORD dynamics. We also fit to these.

PLOT: DYNAMICS UNDER IR, RR AND HL.

Also show dependence on immigration (show that it is low and what happens if it is high), concede that this is a limitation.

1.5.3 Extend methodology (3, 4 and 5 species)

This does not require much since we presented a general framework previously.

Show 3 species dynamics with the two different RR. Conclude which is better.

1.5.4 Results

1.5.5 2 species

Here we compare the results of two species.

For a single IR we look at convergence of all 6 parameters (over the ensemble) - correct signs, correct magnitudes?

Maybe repeat for other IRs and HL.

We then show rate estimates as time series and introduce quality metric for this².

Demonstrate the quality decreases with IR and HL (box plot?). And how estimated parameters respond to the two HL scenarios (refer to previous findings). Hopefully support!

1.6 Discussion

Points referenced in text above, make sure to discuss them!..

- Discuss how this methodology could be used on empirical data...
- Limitations of ODE models (non-spatial, response to debate on FR)
- Possibility of extending to more than two species systems (if this is actually done then change ref in text above)
- discussion of other forms of FR (not H) - or discussed already in intro?
- good GLV fit to LV even with 100 sample points - realistic?

²Still not sure about this

- how good is the method of model fitting. Discuss more computationally expensive options (mentioned in section on Timme method)
- Spatial heterogeneity - we do not explore this here, but acknowledge that it represents are source of error. Gives example in extremis - 2 species only interacting on boundary. Also we know that there is some level of spatial aggregation, especially at low IR - possibly show one plot of this?
- Could introduce prey handling to IBM to create non-linear FR.

BIBLIOGRAPHY

- [1] S. G. SHANDILYA AND M. TIMME, *Inferring network topology from complex dynamics*, New Journal of Physics, 13 (2011), p. 013004.

