Identification of a predator-prey model parameters

Václav Pink

Masaryk University, Faculty of Science, Dept. of Mathematics and Statistics
Kotlářská 2, 611 37 Brno, Czech Republic, E.U.

Abstract. *Predator-prey models lie in a theoretical background of integrated plant protection. The crucial question is whether one is able to identify parameters of such a model on a base of field data.*

The contribution proposes a method of parameter identification for Gause-type predator-prey model. The method is tested on simulated data, in particular, the problem of distinguishing between distinct types of predator functional response (Ivlev and Holling type ones) and of prey growth rate (with limited and unlimited carrying capacity) is tested. Obtained results show that data do not allow one to find such subtilities of the model. Nevertheless, the proposed method is reliable enough for investigation of the system equilibrium and its stability.

Key words and phrases. *Predator-prey, ODE model, parameter identification, equilibrium stability.*
Mathematics Subject Classification. Primary 92D25, 62F10; Secondary 34D05.

I. Introduction

A method of integrated plant protection was introduced into South-Moravian vineyards in early 90's of the last century. One part of a theoretical background for the method was a dynamical model of mites community on leafs — phytophagous mite *Calepitrimerus vitis* (pest) and predatory mite *Typhlodromus pyri* (natural antagonist of the pest) [2]. The proposed model used to yield a prognosis of the pest population density during a vegetation season based on spring time densities of both mite species and, subsequently, it indicated a (non)necessity of a definite chemical treatment of vineyard. The method has been widely used and it works: the vineyards are more healthy, i.e. more productive, and significantly less acaricides are applied, i.e. the production is cheaper and an environment is not exceedingly burdened with chemicals.

But there is a dubitation — various modifications of pest-bioagens (predator-prey) model exist and field data can be fitted not only with one of them. A typical observed population densities of the mentioned mites community and theoretical ones issuing from four different models (for details see below) are plotted on Fig. 1.
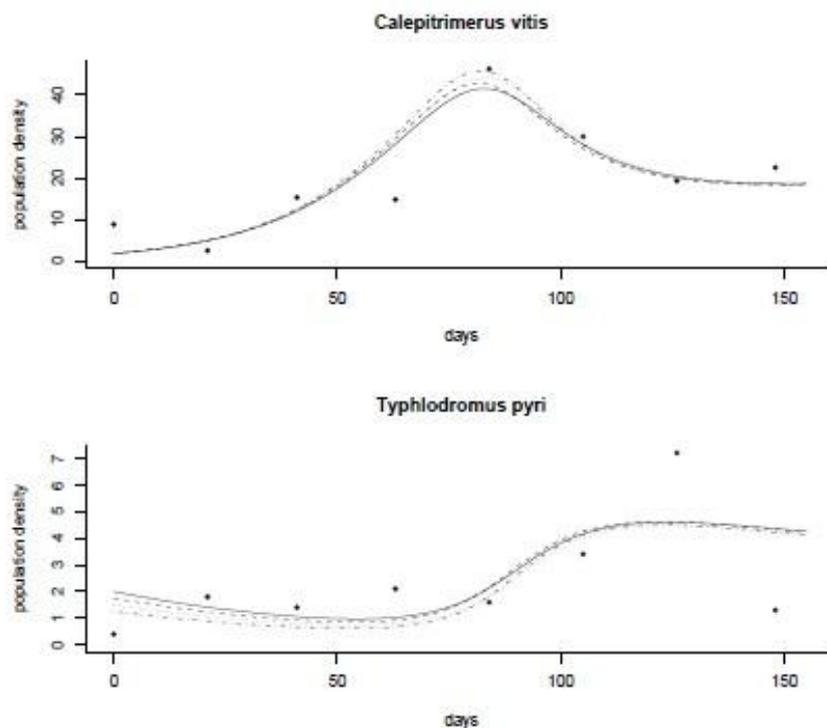
Figure 1. Observed population densities (in individuals per leaf) of phytophagous mite *Cale−pitrimerus vitis* and predatory mite *Typhlodromus pyri* on Dobré Pole in 1989 (bullets) and theoretical evolution of the community based on four types of ODE Gause-type predator-prey models. Solid line — unbounded growth rate and Ivlev-type trophic function, dashed line — unbounded growth rate and Holling-type trophic function, dotted line — bounded growth rate and Ivlev-type trophic function, dotdash line — unbounded growth rate and Holling-type trophic function.

We can see, that different theoretical population density curves are almost indistinguishable. And now, questions arise: Is the model lying in a basis of the mentioned protection method correct? Does a choice of particular model influence a signalizing of a pesticide intervention? Do field data enable us to distinguish subtilities of population dynamics? Is the fact that the method works sufficient for verification of the underlying theory? Answers for these questions is searched in a virtual world. Population dynamics was simulated by four variants of Gause-type predator-prey model and the simulated data was fitted by correct and incorrect models, i.e. by the model used for simulation and by the others.

The subsequent section briefly reminds the predator-prey models, the third one describes a method of parameter identification and the fourth one presents simulations and it summarizes results of parameters identification. Brief discussion of results is added.

II. Model

The well-known and widely used model of predator-prey (or producer-consumer, host-parasite, plant-herbivore etc.) interaction is the Gause-type one:

$$x^{'} = xq(x) - S\varphi(x)y, \qquad y^{'} = -dy + \kappa S\varphi(x)y, \tag{1}$$

where $x = x(t)$ and $y = y(t)$ denote a size (population density, biomass, number of individuals etc.) of prey and predator populations, respectively, q denotes prey growth rate in a surroundings with absent predator (it can depend on size of prey population), d denotes starving predator death rate which is consider to be constant, $\varphi(x)$ is the so called trophic function (rate of satiety of predator population of unit size fed on prey population of size x in a unit time; $\varphi(x) = 0$ – predator is totally starving, $\varphi(x) = 1$ – predator is totally saturated), S denotes the level of satiety of predator (the maximal possible size of prey population destroyed by predator population of unit size in a unit time); consequently, $S\varphi(x)$ represent a size of prey population destroyed by a predator population of unit size in unit time, κ denotes efficiency of predation (fraction of destroyed prey population which is transformed into growth of predator population).

The simplest forms of the prey growth rate function are the constant (Malthus) one: $q(x) = r$ (the prey population does not exhibit neither inter-specific competition nor mutualism) and the linearly decreasing (Verhulst) one: $q(x) = r(1 - x / K)$ (the prey population exhibits an inter-specific competition in an environment with limited resources). Here r denotes the intrinsic growth rate (maximal possible physiological growth rate of population) and K represents the environmental carrying capacity. The constant (i.e. density independent) growth rate can be considered to be a special case of the linearly decreasing one (with $K = \infty$); hence the function q depends on two positive parameters r and K,

$$q(x) = q_{r,K}(x) = r\left(1 - \frac{x}{K}\right), \quad r \in (0,\infty), K \in (0,\infty) \cup \{\infty\}. \tag{2}$$

The trophic function φ should have the following properties:

- $\varphi(0) = 0$ (if none prey is available the predator starves),
- φ is non-decreasing function (if more prey is available the predator is not less satiated),
- $\lim\limits_{x\to\infty} \varphi(x) = 1$ (one predator is able to destroy the prey population up to its saturation).

The widely used functions fulfilling the mentioned conditions are

$$\varphi(x) = 1 - \exp(-ax^s) \quad \text{and} \quad \varphi(x) = \frac{ax^s}{1 + ax^s}, \tag{3}$$

i.e. the Ivlev type and the (Michaelis-Menten-)Holling type trophic function, respectively. The both functions depend on two positive parameters s and a, $\varphi(x) = \varphi_{s,a}(x)$; in the both cases, the parameter s determines

the shape of function (the function φ is concave downward if $s \leq 1$ and it is s-shaped otherwise) and a is a kind of scale parameter characterizing "a speed of convergence" of the function $\varphi_{s,a}$ to unity.

The most important features of the model are:

- if $d < \kappa S$ then there exists the unique interior equilibrium

$$(x^*, y^*) = \left(\varphi^{-1}\left(\frac{d}{\kappa S}\right), \frac{\kappa x^* q(x^*)}{d} \right),$$

-

- if $C_S^* := S y^* \varphi'(x^*) - q(x^*) - x^* q'(x^*) > 0$ then the equilibrium (x^*, y^*) is locally asymptotically stable,

- if $C_O^* := 4 d \kappa S y^* \varphi'(x^*) - (C_S^*)^2 >$ then the equilibrium (x^*, y^*) is focus, i.e. the solution oscillates around it.

For details see e.g. [4] or [1].

III.   Parameters identification

Let $\{(t_i, X_i, Y_i) : i = 0, 1, ..., n\}$, $t_0 < t_1 \cdots < t_n$, be observed data; here t_i represents time of the i-th observation, X_i and Y_i are corresponding sizes of prey and predator population, respectively. An obvious method for identification of model parameters is to evaluate a likelihood L of observed quantities conditioned by model parameters and initial values $x_0 = x(t_0)$, $y_0 = y(t_0)$

$$L = L(\{(t_i, X_i, Y_i)\}, S, d, \kappa, r, K, s, a, x_0, y_0)$$

and to find

$$\mathrm{argmax}\left\{ L(\{(t_i, X_i, Y_i)\}, S, d, \kappa, r, K, s, a, x_0, y_0) : \right.$$
$$\left. S > 0, d > 0, \kappa > 0, r > 0, 0 < K \leq \infty, s > 0, a > 0, x_0 > 0, y_0 > 0 \right\}.$$

But this process is based on some suppositions on probability distribution of data and parameters, it is computationally very demanding since it is necessary to find numerical solution of the system (1) in each iteration step. Moreover, any numerical maximization of the liklihood function depends on estimation of an initial approximation. Therefore, we propose a method of finding a initial estimation of parameters.

Integration of the equations (1) from t_0 to t_i gives

$$x(t_i) = x_0 + r \int_{t_0}^{t_i} x(\tau) \mathrm{d}\tau - b \int_{t_0}^{t_i} (x(\tau))^2 \mathrm{d}\tau - S \int_{t_0}^{t_i} y(\tau) \varphi_{s,a}(x(\tau)) \mathrm{d}\tau,$$

$$y(t_i) = y_0 - d \int_{t_0}^{t_i} y(\tau) \mathrm{d}\tau + \lambda \int_{t_0}^{t_i} y(\tau) \varphi_{s,a}(x(\tau)) \mathrm{d}\tau,$$

(4)

where $\lambda = \kappa S$, $b = r / K$ for $K < \infty$ and $b = 0$ for $K = \infty$. Let f be an integrable function defined on $[t_0, t_n]$ and denote $F_i = f(t_i)$,

$$(IF)_i = \frac{1}{2} \sum_{j=1}^{i} (F_j + F_{j-1})(t_j - t_{j-1}), \quad \text{for } i = 1, 2, ..., n;$$

$(IF)_i$ is the numerical approximation of the integral $\int_{t_0}^{t_i} f(\tau) \mathrm{d}\tau$ by the trapezoid rule. Since $x(t_i) \approx X_i$, $y(t_i) \approx Y_i$ and for s, a fixed all of the integrals appearing in (4) can be approximated by the above formula, the equalities (4) can be rewritten:

$$X_i = x_0 + r(IX)_i - b(IX^2)_i - S(IY\Phi_{s,a})_i + \varepsilon_{1,i},$$
$$Y_i = y_0 - d(IY)_i + \lambda(IY\Phi_{s,a})_i + \varepsilon_{2,i},$$

(5)

where $\varepsilon_{1,i}$, $\varepsilon_{2,i}$ denote cumulative errors of approximation and observation. These equalities can also be considered as a linear regression model. This way, the parameters x_0, r, b, S, y_0, d, λ can be estimated. The standard statistical analysis (F test) can be applied for determination, whether b differs from 0 significantly, i.e. whether the carrying capacity for the prey population is limited or not. This way, a density dependence of prey growth rate might be recognized from data.

Let us denote the estimates depending on the choice of s and a by $x_0^{s,a}$, $\tilde{r}^{s,a}$, $\tilde{b}^{s,a}$, $\tilde{S}^{s,a}$, $y_0^{s,a}$, $\tilde{d}^{s,a}$, $\tilde{\lambda}^{s,a}$, respectively.

The term $(IY\Phi_{s,a})_i$ can be evaluated from the both equalities (5), hence

$$F_i(s,a) := \frac{X_i - x_0^{s,a} - \tilde{r}^{s,a}(IX)_i + \tilde{b}^{s,a}(IX^2)_i}{\tilde{S}^{s,a}} \approx (IY\Phi_{s,a})_i \approx$$

$$\approx \frac{Y_i - y_0^{s,a} + \tilde{d}^{s,a}(IY)_i}{\tilde{\lambda}^{s,a}} =: G_i(s,a)$$

Now, we can put estimates $\hat{s}$, $\hat{a}$ of the parameters s, a to be

$$(\hat{s},\hat{a}) = \arg\min\left\{\sum_{i=1}^{n}(F_i(s,a) - G_i(s,a))^2 : s > 0, a > 0\right\} \tag{6}$$

provided the minimum exists, and then put $x_0 = x_0^{\hat{s},\hat{a}}$, $\hat{r} = \tilde{r}^{\hat{s},\hat{a}}$, $\hat{b} = \tilde{b}^{\hat{s},\hat{a}}$, $\hat{S} = \tilde{S}^{\hat{s},\hat{a}}$, $y_0 = y_0^{\hat{s},\hat{a}}$, $\hat{d} = \tilde{d}^{\hat{s},\hat{a}}$, $\hat{\kappa} = \tilde{\lambda}^{\hat{s},\hat{a}}/\hat{S}$. A model might be considered to be identified if the minimum (6) is (numerically) located and all of the estimated parameters are positive with possible exception of b that can equal to 0.

A quality of data fit can be measured by a distance of numerical solution $x(t_i)$, $y(t_i)$ of the system (1) with the parameters identified from data and of observed values X_i, Y_i, $i = 0,1,...,n$. We can assume that the observed population sizes are random variables from log-normal distribution with mean values equal to theoretical values, i.e. solution of the system (1). (Cf. treatment of popular hare-lynx data mentioned in [3].) Hence, we can express the quality of data fit by euclidian distance of observed and theoretical trajectories in logarithmic phase space:

$$D = \sqrt{\sum_{i=0}^{n}\left[\left(\log\frac{x(t_i)}{X_i}\right)^2 + \left(\log\frac{y(t_i)}{Y_i}\right)^2\right]}; \tag{7}$$

here x and y represents a numerical solution of the system (1) with the estimated parameters.

IV. Simulations and results

We have got a set of time instants $\{0, 21, 41, 63, 84, 105, 126, 148\}$, i.e. the same time instant as the ones in the "typical field data" plotted on Fig 1. For these time instants, four types of data were simulated:

1. growth rate with $K = \infty$, φ of the Ivlev type,
2. growth rate with $K = \infty$, φ of the Holling type,
3. growth rate with $K < \infty$, φ of the Ivlev type,
4. growth rate with $K < \infty$, φ of the Holling type.

The model parameters applied in simulations were $r = 0.51$, $S = 17$, $d = 0.17$, $\kappa = 0.062$, $s = 2$, $a = 3.8 \cdot 10^{-5}$, $K = 500$ (in the cases 3. and 4.). Next, thirty "noised" data sets were generated (samples of log-normal distribution with the mean value equal to the simulated one and with the coefficient of variation equal to 0.1) for each of the simulated data type. The parameters of both models – the ones with Ivlev and Holling functional responses – were identified for each of the generated data sample, stability and equilibrium type of the model with estimated parameters were established and the "goodness of fit" D was evaluated. A noteworthy thing is that the simulated density dependence of prey growth rate ($K < \infty$) was not detected in any of the 60 generated data sets.

Results of data fit are summarized on tab. 1, estimated system equilibria are plotted on fig. 1.

			Identified model	
			φ Ivlev	φ Holling
Simulated data	φ Ivlev $K = \infty$	D median	1.649	1.642
		minimum	0.444	0.418
		maximum	7.130	4.172
		No of identified	24 (80%)	27 (90%)
		correct C_S^*	100%	100%
		correct C_O^*	83%	74%
	φ Holling $K = \infty$	D median	1.378	1.252
		minimum	0.461	0.408
		maximum	5.867	4.338
		No of identified	24 (80%)	22 (73%)
		correct C_S^*	96%	95%
		correct C_O^*	17%	32%
	φ Ivlev $K < \infty$	D median	1.318	1.319
		minimum	0.500	0.466
		maximum	5.663	6.769
		No of identified	25 (83%)	24 (80%)
		correct C_S^*	96%	80%
		correct C_O^*	72%	79%
	φ Holling $K < \infty$	D median	1.531	1.470
		minimum	0.556	0.635
		maximum	8.080	5.219
		No of identified	27 (90%)	24 (80%)
		correct C_S^*	96%	92%
		correct C_O^*	22%	25%

Table 1. Median and extreme values of distance D (7) for simulated and estimated trajectories, number of samples out of 30 generated data sets with successful identification of parameters (i.e. the minimum (6) is located and all of the estimated parameters are positive) and proportions of correctly established stability of the interior equilibrium (x^*, y^*) and the ones of correctly established oscillations around it.

We can see that:

- The distance D of generated data and fitted ones is slightly, but not significantly, smaller for the true model than for the alternative one in the case of $K < \infty$, but this is not true in the case $K = \infty$. Hence, the value of D is not good criterion for a model selection.
- Parameters identification is not more often successful (i.e. the procedure leads to the result when all of the estimated parameters are positive and the minimum (1) is located) for the true model than for the alternative one. Successful parameters identification does not particularly support a model choice.
- Estimated equilibria are not far from the true ones; at least from a point of view of the application indicated.

If identification of parameters method leads to the target then the establishing of equilibrium stability is correct with high probability in spite of the fact that the model used for parameters estimation is different from the one used for data generation. On the contrary, establishing of deterministic oscillations of population sizes around the equilibrium is not very credible.
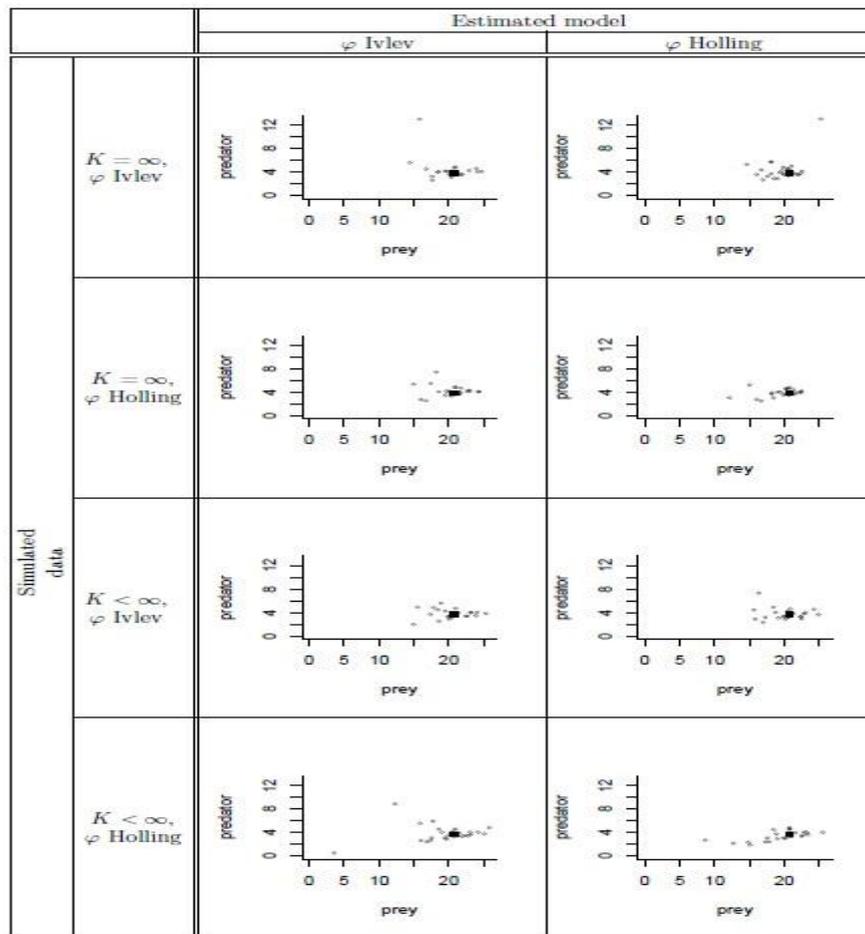
Figure 2. Estimated equilibria of the system (1) for the four generated data and the two ways of parameter identification. The black square denotes the true equilibrium for simulated data.

Shortly saying, a good fit of data does not guarantee that a model is correct; it is not possible to distinguish between various types of trophic function from field data, density dependence of prey growth rate is not detectable, observation does not allow to draw a line between deterministic and stochastic oscillations. On the other hand, finding of system equilibrium an establishing of asymptotic stability of it used to be correct independently on model chosen for data fit.

V.   Discussion

The results obtained show that having field data, it is not possible to distinguish between different models of predator-prey community (at least, if the data are of similar amount as the data samples generated for the presented study; however, it is not possible to have longer or more detailed series of observation for mites community on vineyard which is important from the point of view of the mentioned application). Nevertheless, the determination of community equilibrium and establishing of its stability — the most important feature from the point of view of application — is correct independently on model chosen for data fitting. This result represents one more justification of grapevine protection method utilized for more than ten years on South-Moravian vineyards [2].

References

[1]   Britton, N., F.: *Essential Mathematical Biology.* Springer, London Berlin Heidelberg, 2003.
[2]   Hluchý, M., Pospíšil, Z.: *Use of the predatory mite Typhlodromus pyri Scheuten (Acari: Phytoseiidae) for biological protection of grape vines from phytophagous mites*. In DUSBÁBEK, F., BUKVA, V. (eds.): *Modern Acarology*, Vol. 2. Prague: Academia and SPB Academic Publishing bv, The Hague, 1991, pp. 655–660.
[3]   McCallum, H.: *Population Parameters. Estimation for Ecological Models.* Blackwell Science Ltd., Oxford London Edinburgh Malden Carlton Paris, 2000.
[4]   Svirezhev, Yu., M., Logofet, D., O.: Stability of biological communities [Russian]. Nauka, Moscow, 1978.
.