

# Inference and model determination for Temperature-Driven non-linear Ecological Models: supplemental document

## 1. TABLES WITH SUMMARIES AND POSTERIOR ESTIMATES OF THE STATISTICAL METHODS USED

### A. *Propylea quatuordecimpunctata* dataset

**Table S1.** Algorithmic working time in seconds, for two datasets.

<i>Tetranychus urticae</i>				
Gaussian				
	Bieri	Briere	Analytis	Lactin
HMC	2011	476	9852	5672
ADVI-meanfield	8	1	52	1
ADVI-fullrank	45	6	38	44
Inverse Gamma				
HMC	1375	761	3231	4917
ADVI-meanfield	2	4	15	27
ADVI-fullrank	2	34	14	44
<i>Propylea quatuordecimpunctata</i>				
Gaussian				
	Bieri	Briere	Analytis	Lactin
HMC	4335	20218	4913	1426
ADVI-meanfield	13	7	3	16
ADVI-fullrank	48	11	12	16
Zero Inflated Inverse Gamma				
HMC	1078	321	15291	3073
ADVI-meanfield	41	5	9	10
ADVI-fullrank	52	5	39	18

**Table S2.** BMA weights for the *Tetranychus urticae* data

Gaussian distribution							
	aic_w	dic_w	loocv_w	waic_w	bic_w	elbo_mf	elbo_fr
Bieri	4.9E-17	9.4E-16	2.3E-16	2.4E-16	8.1E-18	3.1E-53	7.3E-45
Briere	1.9E-32	2.3E-31	6.6E-33	6.9E-33	1.1E-31	2.9E-27	1.5E-18
Analytis	1	1	1	1	1	0.001	1
Lactin	1.0E-14	4.9E-14	1.2E-12	1.2E-12	9.0E-18	0.999	3.5E-39
Inverse Gamma distribution							
Bieri	3.0E-43	4.5E-43	1.1E-42	1.0E-42	3.0E-43	1.6E-20	1.0E-200
Briere	6.6E-36	8.1E-36	2.2E-36	2.1E-36	3.8E-35	8.9E-19	1.0E-171
Analytis	1.4E-5	7.1E-6	2.1E-5	2.1E-5	1.9E-6	1	1
Lactin	0.999	0.999	0.999	0.999	1	1.0E-36	2.7E-58

**Table S3.** BMA weights for the *Propylea quatuordecimpunctata* data

Gaussian distribution							
	aic_w	dic_w	loocv_w	waic_w	bic_w	elbo_mf	elbo_fr
Bieri	0.999	0.968	0.991	0.991	1	1	1
Briere	3.3E-34	7.3E-30	5.5E-35	5.5E-35	4.8E-33	5E-203	3E-218
Analytis	1.1E-31	1.5E-32	3.8E-32	3.8E-32	4.0E-31	9.0E-269	1.0E-210
Lactin	5.9E-04	0.032	0.009	0.009	1.1E-05	5.0E-159	3.0E-162
Zero Inflated Inverse Gamma distribution							
Bieri	9.3E-5	0.001	0.001	0.001	3.5E-4	5.1E-53	3.4E-72
Briere	1.7E-38	3.9E-38	7.3E-38	7.2E-38	2.4E-37	1	1.1E-23
Analytis	1	0.999	0.999	0.999	1	3.3E-35	1
Lactin	1.7E-8	2.1E-8	5.8E-8	5.8E-8	6.4E-8	3.0E-114	2.0E-140

**Table S4.** The mean and 95% Cr.I. limits of the BMA estimates of parameters of interest calculated with the use of information criteria score weights assuming the Gaussian and Inverse Gamma distributions respectively for *Tetranychus urticae* data.

Gaussian Model								
		aic_w	dic_w	loocv_w	waic_w	bic_w	elbo_mf	elbo_fr
$T_{min}$	Mean	4.4	4.4	4.4	4.4	4.4	4.4	4.4
	95% Cr.I.	(4.0, 5.3)	(4.0, 5.3)	(4.0, 5.3)	(4.0, 5.3)	(4.0, 5.3)	(4.0, 5.3)	(4.0, 5.3)
$T_{opt}$	Mean	32.9	32.9	32.9	32.9	32.9	32.6	32.9
	95% Cr.I.	(32.7, 33.3)	(32.7, 33.3)	(32.7, 33.3)	(32.7, 33.3)	(32.7, 33.3)	(32.4, 32.8)	(32.7, 33.3)
$T_{max}$	Mean	35.3	35.3	35.3	35.3	35.3	36.9	35.3
	95% Cr.I.	(35.1, 35.5)	(35.1, 35.5)	(35.1, 35.5)	(35.1, 35.5)	(35.1, 35.5)	(36.5, 37.3)	(35.1, 35.5)
dev <sup>†</sup>	Mean	-1750.3	-1750.3	-1750.3	-1750.3	-1750.3	-1693.9	-1750.3
	95% Cr.I.	(-1758.0,-1738.3)	(-1758.0,-1738.3)	(-1758.0,-1738.3)	(-1758.0,-1738.3)	(-1758.0,-1738.3)	(-1703.1,-1679.7)	(-1758.0,-1738.3)
Inverse Gamma Model								
$T_{min}$	Mean	4.2	4.2	4.2	4.2	4.2	4.2	4.2
	95% Cr.I.	(4.0, 4.6)	(4.0, 4.6)	(4.0, 4.6)	(4.0, 4.6)	(4.0, 4.6)	(4.0, 4.6)	(4.0, 4.6)
$T_{opt}$	Mean T <sub>opt</sub>	32.0	32.0	32.0	32.0	32.0	33.6	33.6
	95% Cr.I.	(31.8, 32.2)	(31.8, 32.2)	(31.8, 32.2)	(31.8, 32.2)	(31.8, 32.2)	(33.3, 33.9)	(33.3, 33.9)
$T_{max}$	Mean	38.4	38.4	38.4	38.4	38.4	35.0	35.0
	95% Cr.I.	(38.1, 38.9)	(38.1, 38.9)	(38.1, 38.9)	(38.1, 38.9)	(38.1, 38.9)	(35.0, 35.1)	(35.0, 35.1)
dev <sup>†</sup>	Mean	-1916.3	-1916.3	-1916.3	-1916.3	-1916.3	-1894.0	-1894.0
	95% Cr.I.	(-1920.6,-1908.3)	(-1920.6,-1908.3)	(-1920.6,-1908.3)	(-1920.6,-1908.3)	(-1920.6,-1908.3)	(-1899.2,-1885.2)	(-1899.2,-1885.2)

† deviance of the model given the data.

**Table S5.** The mean and 95% Cr.I. limits of the BMA estimates of parameters of interest calculated with the use of information criteria score weights assuming the Gaussian and Inverse Gamma distributions respectively for the *Propylea quatuordecimpunctata* data.

Gaussian distribution								
		aic_w	dic_w	loocv_w	waic_w	bic_w	elbo_mf	elbo_fr
$T_{min}$	Mean	10.6	10.6	10.6	10.6	10.6	10.6	10.6
	95% Cr.I.	(9.7, 11.4)	(9.7, 11.4)	(9.7, 11.4)	(9.7, 11.4)	(9.7, 11.4)	(9.7, 11.4)	(9.7, 11.4)
$T_{opt}$	Mean	32.6	32.5	32.6	32.6	32.6	32.6	32.6
	95% Cr.I.	(32.0, 33.5)	(31.9, 33.4)	(32.0, 33.5)	(32.0, 33.5)	(32.0, 33.5)	(32.0, 33.5)	(32.0, 33.5)
$T_{max}$	Mean	35.0	35.0	35.0	35.0	35.0	35.0	35.0
	95% Cr.I.	(34.98, 35.0)	(34.98, 35.0)	(34.98, 35.0)	(34.98, 35.0)	(34.98, 35.0)	(34.98, 35.0)	(34.98, 35.0)
dev <sup>†</sup>	Mean	-818.8	-818.5	-818.7	-818.7	-818.8	-818.8	-818.8
	95% Cr.I.	(-833.1, -800.8)	(-832.4, -801.1)	(-832.9, -800.9)	(-832.9, -800.9)	(-833.1, -800.8)	(-833.1, -800.8)	(-833.1, -800.8)
Inverse Gamma distribution								
$T_{min}$	Mean	5.0	5.0	5.0	5.0	5.0	11.1	5.0
	95% Cr.I.	(4.0, 7.0)	(4.0, 7.0)	(4.0, 7.0)	(4.0, 7.0)	(4.0, 7.0)	(10.3, 11.9)	(4.0, 7.0)
$T_{opt}$	Mean	33.5	33.5	33.5	33.5	33.5	29.3	33.5
	95% Cr.I.	(32.3, 34.9)	(32.3, 34.9)	(32.3, 34.9)	(32.3, 34.9)	(32.3, 34.9)	(29.2, 29.5)	(32.3, 34.9)
$T_{max}$	Mean	33.6	33.6	33.6	33.6	33.6	35.0	33.6
	95% Cr.I.	(32.5, 34.9)	(32.5, 34.9)	(32.5, 34.9)	(32.5, 34.9)	(32.5, 34.9)	(34.8, 35.0)	(32.5, 34.9)
dev <sup>†</sup>	Mean	-736.2	-736.2	-736.2	-736.2	-736.2	-563.8	-736.2
	95% Cr.I.	(-742.1, -729.1)	(-742.1, -729.1)	(-742.0, -729.1)	(-742.0, -729.1)	(-742.1, -729.1)	(-568.0, -555.9)	(-742.1, -729.1)

† deviance of the model given the data.

**Table S6.** Posterior summaries for the four models using the Gaussian distribution for the *Tetranychus urticae* data. In each column we report the HMC, the ADVI-Mean field and ADVI-Full rank estimates respectively.

	$T_{min}$			neff <sup>†</sup>	$T_{opt}$			neff <sup>‡</sup>
Bieri	10.5 (10.2, 10.8)	9.8 (9.5, 10.2)	9.8 (9.2, 10.3)	15089	33.0 (32.7, 33.4)	158.6 (144.4, 165.4)	183.8 (76.4, 368.5)	8164
Briere	9.3 (8.6, 9.9)	9.3 (9.1, 9.5)	9.3 (8.6, 9.9)	8690	33.1 (32.5, 33.7)	33.0 (32.8, 33.2)	33.1 (32.4, 33.7)	8028
Analytis	4.4 (4.0, 5.3)	4.3 (4.2, 4.4)	5.3 (4.8, 6.1)	8757	32.9 (32.7, 33.3)	33.0 (32.9, 33.1)	33.2 (32.6, 33.7)	41
Lactin	10.4 (10.1, 10.7)	-3.9 (-5.8, -1.9)	-0.7 (-4.0, 3.0)	11526	32.6 (32.4, 32.8)	32.2 (32.0, 32.3)	32.0 (31.8, 32.2)	9884
	$T_{max}$			neff <sup>†</sup>	dev <sup>†</sup>			neff <sup>‡</sup>
Bieri	36.3 (35.9, 36.8)	164.9 (161, 169)	190.6 (80, 376.1)	7561	-1677.1 (-1683.2, -1666.6)	-1461.0 (-1467.7, -1443.6)	-1461.7 (-1467.7, -1456.3)	10087
Briere	40.0 (39.3, 40.9)	39.9 (39.6, 40.1)	40.0 (39.1, 40.9)	7562	-1602.2 (-1606.8, -1593.4)	-1593.2 (-1605.7, -1569.1)	-1598.5 (-1606.3, -1583.6)	9618
Analytis	35.3 (35.1, 35.5)	35.2 (35.2, 35.3)	35.2 (35.1, 35.4)	8246	-1750.3 (-1758.0, -1738.3)	-1744.7 (-1755.6, -1724)	-1709.1 (-1741.9, -1606.9)	10954
Lactin	36.9 (36.6, 37.3)	38.3 (38.1, 38.5)	38.2 (37.9, 38.5)	7797	-1693.9 (-1703, -1679.6)	-1764.0 (-1780.3, -1738.1)	-1776.2 (-1784.9, -1762)	8227

<sup>†</sup> deviance of the model given the data,

<sup>‡</sup> effective sample size.

**Table S7.** Posterior summaries for the four models using the Gaussian distribution for the Propylea Coccinellidae data. In each column we report the HMC, the ADVI-Mean field and ADVI-Full rank estimates respectively.

		$T_{min}$			neff <sup>†</sup>	$T_{opt}$			neff <sup>‡</sup>
Mean 95% Cr. I.	Bieri	10.6 (9.7, 11.4)	10.7 (10.6, 10.8)	10.7 (10, 11.4)	7906	32.6 (32, 33.5)	32.2 (32, 32.5)	32.2 (31.8, 32.5)	7620
	Briere	13.1 (12.4, 13.8)	13.3 (12.9, 13.8)	8.2 (1.8, 15.6)	4	29.7 (29.6, 29.8)	29.7 (29.5, 29.8)	27.3 (11.7, 30.1)	4
	Analytis	6.3 (4.1, 10.6)	21.5 (7.7, 55.2)	6.9 (5.4, 9.3)	3742	33.2 (32.1, 34.8)	39.2 (18.1, 89.1)	33.2 (31.7, 37.7)	4435
	Lactin	-154.4 (-234.7, -119.3)	-146.2 (-154.3, -138.1)	-159.2 (-243.1, -119.0)	7560	31.2 (30.9, 31.3)	31.1 (31.1, 31.2)	31.1 (30.9, 31.3)	8024
		$T_{max}$			neff <sup>†</sup>	dev <sup>†</sup>			neff <sup>‡</sup>
Mean 95% Cr. I.	Bieri	35.0 (34.98, 35.02)	33.8 (33.6, 33.9)	33.7 (33.5, 33.9)	7194	-818.7 (-833.1, -800.6)	-672.7 (-685.1, -657.4)	-672.5 (-686.2, -657.1)	11673
	Briere	35.0 (35.0, 35.01)	35.0 (34.8, 35)	32.9 (14.3, 35)	113	-660.5 (-667.0, -649.7)	-535.8 (-543.2, -520.5)	-429.7 (-535.0, -402.1)	14478
	Analytis	33.6 (32.5, 34.9)	42.5 (32.7, 94.5)	36.1 (4.6, 133.2)	9370	-674.1 (-690.2, -666.3)	-160.0 (-261.3, -98.2)	-491.9 (-595.9, -293.6)	5164
	Lactin	35.0 (35, 35.04)	35.0 (35.0, 35.1)	35.0 (35.0, 35.1)	21622	-809.8 (-823.7, -792.2)	-804.7 (-820.5, -784.8)	-778.5 (-807.5, -702.6)	12147

<sup>†</sup> deviance of the model given the data,

<sup>‡</sup> effective sample size.

**Table S8.** Model selection criteria for the eight models applied to the *Propylea Coccinellidae* data

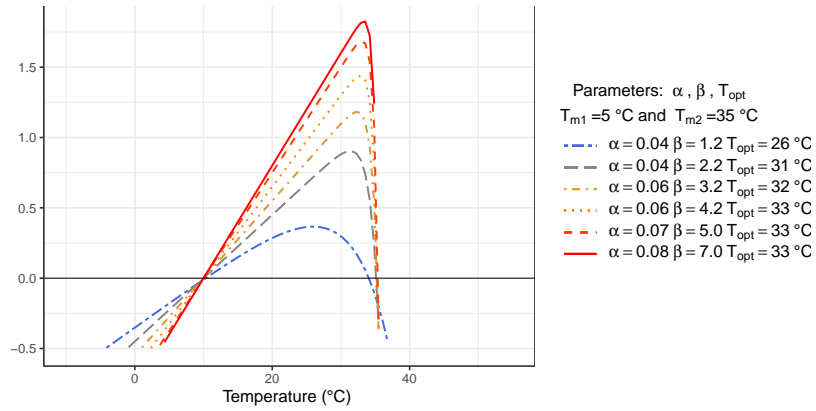
		AIC	DIC	LooCV	WAIC	BIC	$\log(P_yIS)^{\dagger}$ (se)	$\log(P_yPP)^{\ddagger}$ (se)	$\log(P_yBS)^{\S}$ (se)
Gaussian model	Bieri	-804.7	-783.6	-816.1	-816.1	-786.1	419.9 (19.6)	433.4 (18.0)	462.4 (9.2)
	Briere	-650.5	-649.5	-658.3	-658.3	-637.3	299.2 (24.6)	235.4 (6.7)	231.4 (1.2)
	Analytis	-662.1	-637.1	-671.4	-671.4	-646.1	283.7 (24.0)	312.2 (10.0)	274.2 (33.7)
	Lactin	-789.8	-776.8	-806.7	-806.7	-763.3	390.1 (33.5)	328.1 (23.0)	334.7 (1.7)
Inverse Gamma model	Bieri	-716.1	-719.3	-719.3	-719.3	-702.8	334.5 (25.9)	342.9 (31.5)	333.2 (5.0)
	Briere	-560.8	-561.4	-561.4	-561.4	-550.2	268.2 (23.2)	265.7 (4.6)	251.7 (1.9)
	Analytis	-734.7	-732.4	-732.4	-732.4	-718.8	340.8 (26.7)	371.3 (14.5)	335 (31.2)
	Lactin	-698.9	-699.1	-699.1	-699.1	-685.6	329.2 (19.5)	329.3 (11.8)	313.7 (18.8)

$\dagger \log(P_yIS)$  denotes the logarithm of estimated marginal likelihood via Importance sampling,

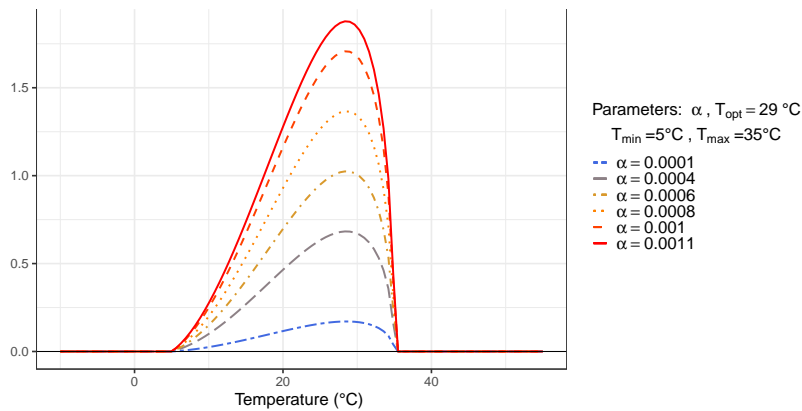
$\ddagger \log(P_yPP)$  denotes the logarithm of estimated marginal likelihood via Power posterior,

$\S \log(P_yBS)$  denotes the logarithm of estimated marginal likelihood via Bridge sampling.

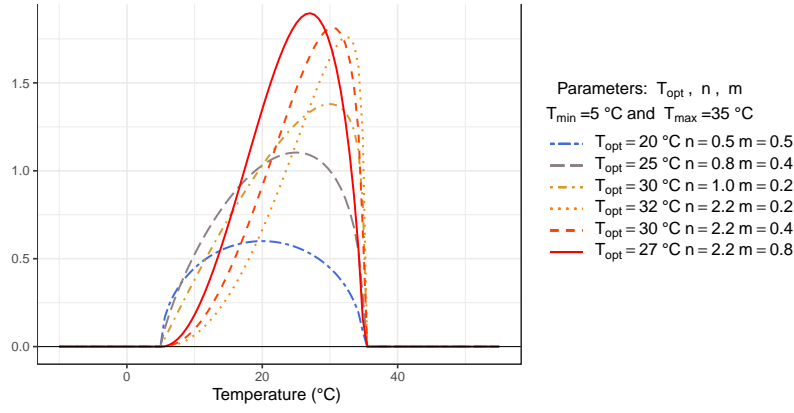
## 2. ECOLOGICAL MODELS CURVES AND POSTERIOR PREDICTION GRAPHS FOR BOTH DATASETS



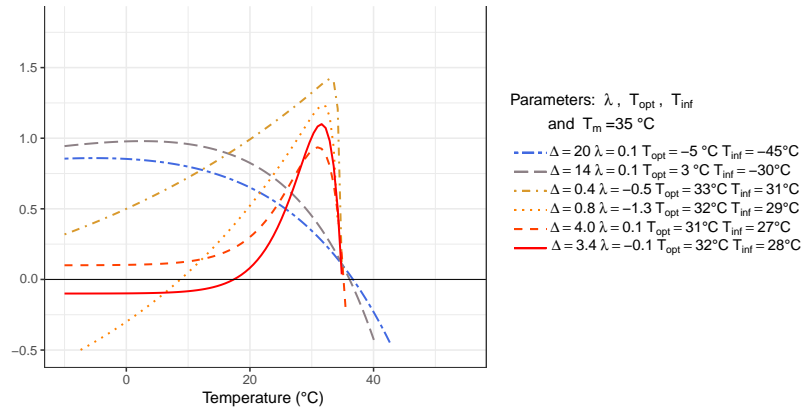
**Fig. S1.** Bieri developmental rates



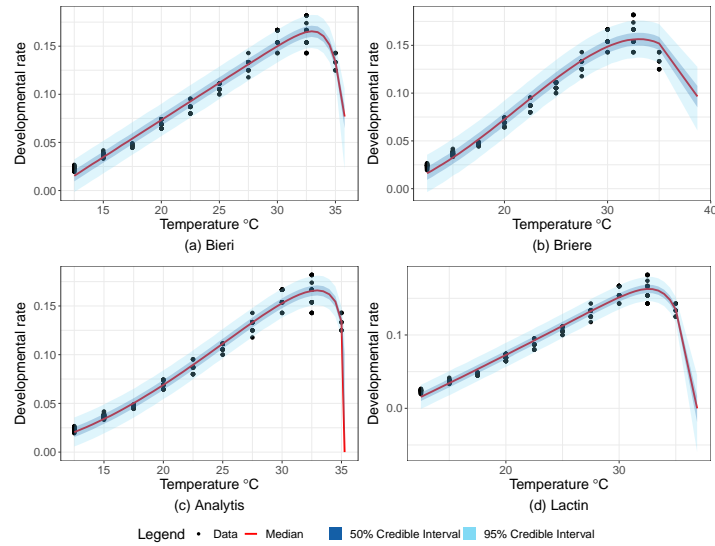
**Fig. S2.** Briere developmental rates



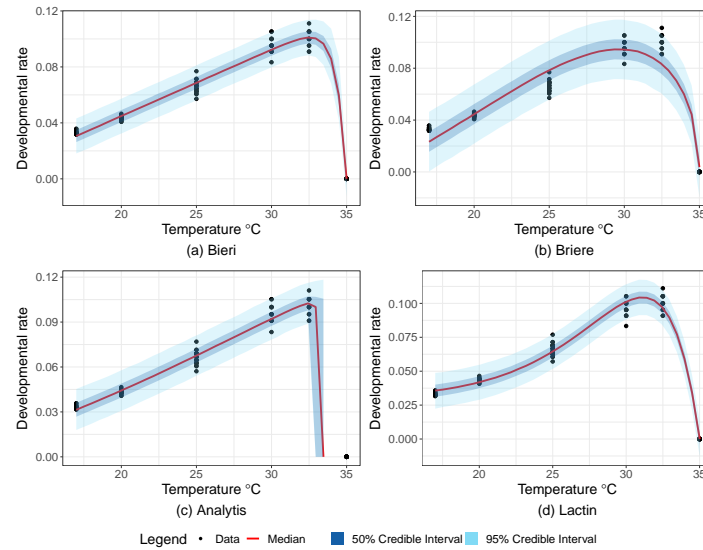
**Fig. S3.** Analytis developmental rates



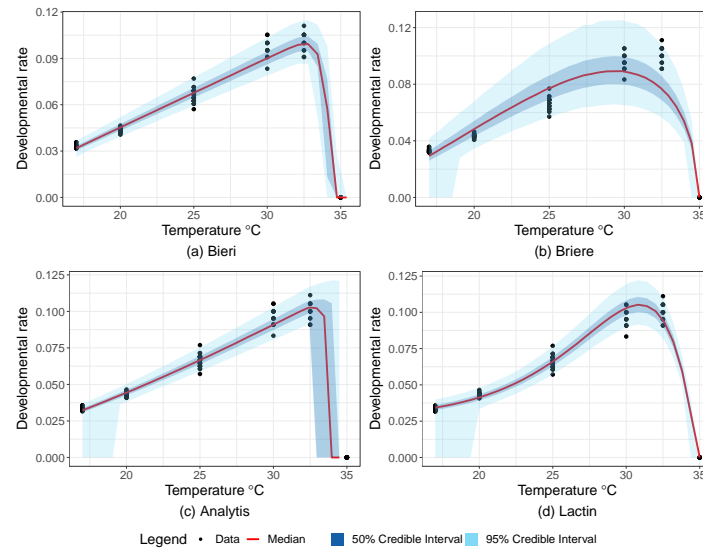
**Fig. S4.** Lactin developmental rates



**Fig. S5.** Predicted posteriors versus *Tetranychus urticae* data using Gaussian distribution



**Fig. S6.** Predicted posteriors versus *Propylea Coccinellidae* data using Gaussian distribution



**Fig. S7.** Predicted posteriors versus *Propylea Coccinellidae* data using Zero Inflated Inverse Gamma distribution

### 3. BAYESIAN INFERENCE

In the current section, we provide details of statistical model specifications in the Bayesian framework. Specifically, we provide a brief description of the most important features of Stan’s implementation of HMC and VB so the reader can get familiar with the tools that Stan is based on. We then provide model selection and model averaging techniques in order to compare the different ecological models and to explore and interpret the parameters of interest combining predictions from all the four of them.

#### A. HMC and VB techniques

The HMC method is a Monte Carlo technique that uses Hamiltonian dynamics in order not only to explore efficiently the target distribution but also to propose distant samples in the parameter space that do not exclusively depend on the current state of the Markov chain like considered in previous MCMC methodology [1]. In this way, many performance challenges are tackled like either the low convergence due to the fact that the parameter space with high posterior support is not reached or the poor exploration of the target distribution due to its multi-modality or its shape irregularities. The existence of Hamiltonian dynamics in the system of the joint density mass function allows the preservation of volume and hence adequate trajectories can be used to define complex mappings of the parameter state space without the need to account for cumbersome Jacobian calculations [2]. Thus, by carefully designing automated trajectory realizations in the Hamiltonian dynamics system, the Stan team managed to create an augmented software called STAN [3], which materializes HMC sampling for the parameters of interest.

Moreover, independently, Automatic Differentiation Variational Inference (ADVI) technique is referred to the machine learning field [4]. The latter is a VB method and posterior target distributions are approximated by choosing the closest distribution to a parametric family of tractable distributions like the exponential family via optimization. In order to achieve this point-wise estimations of the parameters of the family distribution are estimated so that the Kullback–Leibler ‘KL’ divergence function is minimized. Specifically, since the KL divergence is intractable the Evidence Lower Bound is maximized instead [4].

#### B. Model selection and model averaging

In case there are  $m$  models  $(\mathcal{M}_1, \dots, \mathcal{M}_m)$  under consideration, the posterior probability of the suitability of the  $i^{th}$  model given the data  $\mathbf{y}$ , is given by

$$p(\mathcal{M}_i|\mathbf{y}) = \frac{p(\mathbf{y}|\mathcal{M}_i)p(\mathcal{M}_i)}{\sum_{k=1}^m p(\mathbf{y}|\mathcal{M}_k)p(\mathcal{M}_k)} \quad (S1)$$

where  $p(\mathcal{M}_i)$  expresses the prior belief for the  $i^{th}$  model, while the  $p(\mathbf{y}|\mathcal{M}_i)$  is the model evidence also called ‘marginal likelihood’ and it can be interpreted as the likelihood over the space of models, marginalizing out the parameters of the  $i^{th}$  model. The ratio of the marginal likelihoods between two models  $p(\mathbf{y}|\mathcal{M}_i)/p(\mathbf{y}|\mathcal{M}_j)$  is called Bayes factor and is the posterior odds of the null hypothesis that the  $i_{th}$  model fits better the data than the  $j_{th}$  model does when the prior probability of the null is one-half [5]. The Bayes factor is used in order to give evidence for the most probable model given the data, when comparing two alternative models [5].

In the case of Bayesian model averaging, the model selection uncertainty is taking into account in statistical inference. The joint posterior  $p(\mathcal{M}_i, \boldsymbol{\theta}_i|\mathbf{y})$  of the  $i^{th}$  model with vector of parameters  $\boldsymbol{\theta}_i$ , using the Baye’s rule, is proportional to the product of the likelihood of the  $i^{th}$  model times the prior distribution of the parameters  $p(\boldsymbol{\theta}_i)$  times the prior distribution  $p(\mathcal{M}_i)$  (that expresses our uncertainty of the  $i^{th}$  model)

$$p(\mathcal{M}_i, \boldsymbol{\theta}_i|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}_i, \mathcal{M}_i) \cdot p(\boldsymbol{\theta}_i) \cdot p(\mathcal{M}_i). \quad (S2)$$



The uncertainty of the  $i^{th}$  model given the data can then be re-expressed via the posterior probability  $p(\mathcal{M}_i|\mathbf{y})$  defined by the ratio in (S1), in case of existence of multiple models. The posteriors of the models can be thought as weights that are critical to the Bayesian model averaging as they can be used to extract useful weighted statistics from the data distribution while at the same time taking into account model uncertainties. Estimation of model parameters and model uncertainties can be achieved either by directly sampling from the joint posterior (S2) or by approximating the marginal likelihood of each model independently and, accordingly, by controlling the outcomes with a view to formulating proper weights and proceeding with the calculation of the averaged statistics. For the former case, techniques like the reversible jump MCMC [6, 7] and variable selection samples [8–10] are used. On the other hand, for the later case, techniques of marginal likelihood approximations via thermodynamic integration [11], bridge sampling [12], importance sampling [13] or via information- criteria perspective like in [5] are used.

### B.1. Information criteria

The criteria used in the current work are the Akaike information criterion ‘AIC’, the Bayesian information criterion ‘BIC’, the Deviance information criterion ‘DIC’, the Watanabe–Akaike information criterion ‘WAIC’ and the Leave-one-out cross-validation criterion ‘Loocv’. Briefly, these criteria provide an approximation of the expected log predictive density for new-coming data while correcting bias from data usage. In particular AIC [14] is defined as the difference

$$AIC(\mathcal{M}_i) = -2 \log p(\mathbf{y}|\hat{\boldsymbol{\theta}}_i) + 2k_i$$

where  $\hat{\boldsymbol{\theta}}_i$  is the maximum likelihood estimate ‘MLE’ of the  $k_i$  parameters of the  $i^{th}$  model. Similarly, BIC [15] is defined as the difference

$$BIC(\mathcal{M}_i) = -2 \log p(\mathbf{y}|\hat{\boldsymbol{\theta}}_i) + k_i \cdot \log(n)$$

where  $n$  is the sample size. In addition, DIC [16] is defined as the following difference:

$$DIC(\mathcal{M}_i) = -2 \log p(\mathbf{y}|\hat{\boldsymbol{\theta}}_i) + 2p_{DIC}$$

where  $\hat{\boldsymbol{\theta}}_i$  is the posterior mean of the parameters of the  $i^{th}$  model, whereas  $p_{DIC}$  is the effective number of parameters and it is evaluated following [16, 17] by either

$$p_{DIC_1} = E_{\boldsymbol{\theta}_i|\mathbf{y}} \{-2 \log p(\mathbf{y}|\boldsymbol{\theta}_i)\} + 2 \log \left\{ p(\mathbf{y}|\hat{\boldsymbol{\theta}}_i) \right\},$$

or

$$p_{DIC_2} = \frac{Var_{\boldsymbol{\theta}_i|\mathbf{y}} \{\log p(\mathbf{y}|\boldsymbol{\theta}_i)\}}{2},$$

where  $E_{\boldsymbol{\theta}_i|\mathbf{y}} \{\log p(\mathbf{y}|\boldsymbol{\theta}_i)\}$  is an expectation over the posterior density of  $\boldsymbol{\theta}_i$ , whereas  $Var_{\boldsymbol{\theta}_i|\mathbf{y}} \{\log p(\mathbf{y}|\boldsymbol{\theta}_i)\}$  is the variance of the log likelihood of observed data  $\mathbf{y}$ , over the posterior density of  $\boldsymbol{\theta}_i$ . Furthermore, WAIC [18] is defined as the following difference:

$$WAIC(\mathcal{M}_i) = -2 \sum_{j=1}^N \log E_{\boldsymbol{\theta}_i} \{p(y_j|\boldsymbol{\theta}_i)\} + 2p_{WAIC}$$

where  $E_{\boldsymbol{\theta}_i} \{p(y_j|\boldsymbol{\theta}_i)\}$  is the expectation of the probability at  $y_j$  data point over the posterior distribution of the parameters of the  $i^{th}$  model, whereas  $p_{WAIC}$  is the effective number of parameters and it is evaluated following [17] by either

$$p_{WAIC_1} = 2 \sum_{j=1}^N \left[ \log E_{p(\boldsymbol{\theta}_i|\mathbf{y})} \{p(y_j|\boldsymbol{\theta}_i)\} - E_{p(\boldsymbol{\theta}_i|\mathbf{y})} \{\log p(y_j|\boldsymbol{\theta}_i)\} \right]$$

or

$$p_{WAIC_2} = \sum_{j=1}^N \text{var}_{p(\theta_i|y)} \{ \log p(y_j|\theta_i) \}$$

where  $E_{p(\theta_i|y)} \{ \log p(y_j|\theta_i) \}$  is the expectation of the logarithm at value  $y_j$  over the posterior density of  $\theta_i$ , whereas  $\text{var}_{p(\theta_i|y)} \{ \log p(y_j|\theta_i) \}$  is the variance of the log posterior density of the observed data  $y_j$ , over the posterior density of  $\theta_i$ .

Furthermore, LooCV [17] is defined as the following difference:

$$\text{LooCV}(\mathcal{M}_i) = -2 \sum_{j=1}^N \log E_{\theta_i^{-j}} \{ p(y_j|\theta_i^{-j}) \} - 2\beta_{\text{LooCV}}$$

where  $E_{\theta_i^{-j}} \{ p(y_j|\theta_i^{-j}) \}$  is the expectation of the probability at  $y_j$  data point over the posterior distribution of the parameters of the  $i^{\text{th}}$  model. The posterior distribution  $p(\theta_i^{-j}|\mathbf{y}_{-j})$  is sampled considering a partition of the data, leaving one data value ( $y_j$ ) out of the original sample. The  $\beta_{\text{LooCV}}$  is a bias correction of the measure and it is evaluated following [17] by

$$\beta_{\text{LooCV}} = \sum_{j=1}^N \log E_{\theta_i} \{ p(y_j|\theta_i) \} - \frac{1}{N} \sum_{\kappa=1}^N \sum_{j=1}^N \log E_{\theta_i^{-\kappa}} \{ p(y_j|\theta_i^{-\kappa}) \}$$

where  $E_{\theta_i^{-\kappa}} \{ p(y_j|\theta_i^{-\kappa}) \}$  is the expectation of the probability at  $y_j$  data point over the posterior distribution of the parameters of the  $i^{\text{th}}$  model leaving out the  $\kappa^{\text{th}}$  observation.

## B.2. Marginal likelihood estimation techniques

The marginal likelihood can be viewed as a normalizing constant  $z_i = p(\mathbf{y}|\mathcal{M}_i)$  of the density  $q(\theta_i|\mathbf{y}) = p(\mathbf{y}|\theta_i) \cdot p(\theta_i)$  within the  $i^{\text{th}}$  ecological model that includes parameters  $\theta_i$ . In the general scheme of comparing the two densities  $q_0$  and  $q_1$  of interest, as in the case of the Bayes factor of two models or in the case model's prior and posterior, a general path from  $q_0$  to  $q_1$  can be created according to [19] using a class of densities  $p(\theta_i|\mathbf{y}, t)$  on the same space indexed by the continuous auxiliary variable say  $t \in [0, 1]$ . A key formula that links the corresponding normalizing constant  $z(t)$  and the unnormalized density  $q(\theta_i|\mathbf{y}, t)$  that correspond to the sampling distribution  $p(\theta_i|\mathbf{y}, t)$  is given by:

$$\frac{d}{dt} \log z(t) = \int \frac{1}{z(t)} \frac{d}{dt} q(\theta_i|\mathbf{y}, t) p(\theta_i|t) d\theta_i = E_t \left\{ \frac{d}{dt} \log q(\theta_i|\mathbf{y}, t) \right\}, \quad (\text{S3})$$

where the expectation is with respect to the sampling distribution  $p(\theta_i|\mathbf{y}, t)$ .

In addition, another key formula of estimating a ratio of normalizing constants has been of great interest such as in computing likelihood ratios in hypothesis testing or in computational physics in estimating free energy differences, or in computing the Bayes factor in Bayesian framework [12]. The general formula is as follows:

$$\frac{z_1}{z_0} = \frac{p(\mathbf{y}|\theta, t=1)}{p(\mathbf{y}|\theta, t=0)} = \frac{E_0 \{ h(\theta) \cdot q(\theta|\mathbf{y}, t=1) \}}{E_1 \{ h(\theta) \cdot q(\theta|\mathbf{y}, t=0) \}} \quad (\text{S4})$$

where  $E_0$  and  $E_1$  expectations are with respect to posterior distribution densities  $p(\theta|\mathbf{y}, t=0)$  and  $p(\theta|\mathbf{y}, t=1)$  respectively, whereas the bridge function  $h(\theta)$  is defined and overlapped by the common support of the former densities.

Using general formulas (S3) and (S4), several marginal probability evaluation schemes of the  $i^{\text{th}}$  model are derived [19]. The power posterior sampling [11], the importance sampling [13] and the bridge sampling [12, 20] techniques are used for the current work.

In the power posterior case, formula (S3) is integrated with respect to variable  $t$  and  $q(\theta_i|\mathbf{y}, t)$  is substituted with density  $p(\theta_i|\mathbf{y})^t p(\theta_i)$ . The marginal likelihood  $z_i = p(\mathbf{y}|\mathcal{M}_i)$  is derived from logarithmic scale by the equation:

$$\log \{p(\mathbf{y}|\mathcal{M}_i)\} = \int_0^1 E_{\theta_i|\mathbf{y}, t} \{\log p(\mathbf{y}|\theta_i)\} dt \quad (\text{S5})$$

where expectation  $E_{\theta_i|\mathbf{y}, t}$  is taken with respect to the density  $p(\theta_i|\mathbf{y})^t p(\theta_i)$  which is defined as the power posterior at temperature  $t$  [11].

Additionally, the standard error  $se_i$  for the  $i$ th model estimator (S5), as shown in section 5 of the appendix is approximated by:

$$s\hat{e}_i = \sqrt{\frac{(t_2 - t_1)^2}{2} s_1^2 + \sum_{k=2}^{n-1} \frac{(t_k - t_{k-1})^2}{2} s_k^2 + \frac{(t_n - t_{n-1})^2}{2} s_n^2},$$

where  $t_k$  is the time after discretization  $0 = t_0 < t_1 < t_k < t_n = 1$  and  $s_k$  is the standard error of the corresponding estimation  $\log \{p(\mathbf{y}|\mathcal{M}_i)\}$  given in (S5).

In the case of importance sampling, the marginal likelihood is assessed by introducing the proper density function  $g$ . After sampling from the proposed density function  $g$ , the marginal likelihood is calculated as with respect to  $g$  as:

$$p(\mathbf{y}|\mathcal{M}_i) = E_g \left\{ \frac{q(\theta_i|\mathbf{y})}{g(\theta_i)} \right\}$$

Following [13], we use the density  $q(\theta_i|\mathbf{y})$  equal to  $p(\mathbf{y}|\theta_i, \phi_i) \cdot p(\theta_i, \phi_i)$  and the auxiliary importance function  $g$  used is as follows:

$$g(\theta_i) = g(\theta_i, \phi_i) = p(\theta_i|\mathbf{y})p(\phi_i|\mathbf{y}), \quad (\text{S6})$$

where  $(\theta_i, \phi_i)$  are the parameters of the  $i^{th}$  model divided into two blocks  $\theta_i$  and  $\phi_i$  which may or may not be independent. The right hand side of (S6) is the product of the marginal posterior distributions of the block. Thus, the marginal probability which gives the target value is given as follows:

$$p(\mathbf{y}|\mathcal{M}_i) = \iint \frac{p(\mathbf{y}, \theta_i, \phi_i)}{g(\theta_i, \phi_i)} g(\theta_i, \phi_i) d(\theta_i, \phi_i) = E_g \left\{ \frac{p(\mathbf{y}, \theta_i, \phi_i)}{g(\theta_i, \phi_i)} \right\} \quad (\text{S7})$$

The standard error  $se_i$  of (S7) as shown in section 6 of the appendix is:

$$s\hat{e}_i = \sqrt{\frac{1}{K} \sum_{j=1}^K \left\{ \frac{p(\mathbf{y}|\theta^j, \phi^j) \cdot p(\theta^j, \phi^j)}{g(\theta^j)} - \hat{z}_i \right\}^2},$$

where  $\hat{z}_i$  is the estimation of the corresponding marginal probability (in the same form of (S7)), while  $(\theta^j, \phi^j)$  are draws  $j = 1, 2, \dots, K$  from the importance function in (S6).

Additionally, using an alternative version of (S4) in [12, 20, 21] the marginal likelihood of a single model is evaluated using bridge sampling by the formula:

$$z_i = p(\mathbf{y}|\mathcal{M}_i) = \frac{E_g \{h(\theta_i) \cdot q(\theta_i|\mathbf{y})\}}{E_p \{h(\theta_i) \cdot g(\theta_i)\}}, \quad (\text{S8})$$

where  $E_g$  and  $E_p$  are the expectations with respect to  $g(\theta_i)$  a so-called proposal distribution and to  $p(\theta_i|\mathbf{y})$  the  $i^{th}$  model posterior distribution respectively.

The bridge function  $h(\theta_i)$  is selected to minimize the relative mean-squared error of (S4). Following [12] the bridge function is specified by:

$$h(\theta_i) = C \cdot \frac{1}{s_1 \cdot q(\theta_i|\mathbf{y}) + s_2 \cdot p(\mathbf{y}) \cdot g(\theta_i)}, \quad (\text{S9})$$

where  $s_1 = N_1 / (N_1 + N_2)$ ,  $s_2 = N_2 / (N_1 + N_2)$  and  $\mathcal{C}$  is a constant.  $N_1$  is the sample size from the posterior and  $N_2$  is the sample size from  $g(\theta_i)$ .

The optimal bridge function in (S9) includes the marginal likelihood under-assessment so that it cannot be evaluated directly. For this purpose the iterative method suggested by [12] and applied in [22] in R software [23] is used. The alternatives used in place of distribution  $g$  is either a multivariate normal distribution with mean vector and covariance matrix that match the respective posterior samples quantities or a standard multivariate normal distribution in combination with a warped posterior distribution of which the first three moments correspond to [22].

Moreover, following [21] the relative mean square error  $RE_i^2 = \frac{E\{\hat{z}_i - z_i\}^2}{z_i^2}$  of (S8) is evaluated by the formula:

$$\widehat{RE}_i^2 = \frac{1}{N_2} \frac{V_g\{f_1(\theta_i)\}}{E_g^2\{f_1(\theta_i)\}} + \frac{\rho_{f_2}(0)}{N_1} \frac{V_p\{f_2(\theta_i)\}}{E_p^2\{f_2(\theta_i)\}}, \quad (\text{S10})$$

where  $f_1(\theta_i) = \frac{q(\theta_i|y)}{s_1 \cdot q(\theta_i|y) + s_2 \cdot g(\theta_i)}$ ,  $f_2(\theta_i) = \frac{g(\theta_i)}{s_1 \cdot q(\theta_i|y) + s_2 \cdot g(\theta_i)}$ ,  $V_g(f_1(\theta_i)) = \int \{f_1(\theta_i) - E(f_1(\theta_i))\}^2 g(\theta_i) d\theta$  is the variance of  $f_1(\theta_i)$  with respect to  $g(\theta_i)$ . The term  $\rho_{f_2}(0)$  in (S10) corresponds to the normalized spectral density of the auto-correlated process  $f_2(\theta_i)$  at the frequency 0.

Following [22] the square root of  $RE^2$  can be interpreted as coefficient of variation provided that the bridge sampling estimator  $\hat{z}_i$  is unbiased. Then the standard error  $se_i$  of the bridge estimator is evaluated by the product  $\widehat{se}_i = \widehat{RE} \cdot E(\hat{z}_i)$

### B.3. BMA weights

We can derive a weighted prediction  $\tilde{y}$  over the  $m$  different models  $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_m$  predictions  $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m$  by imposing appropriate weights  $w_1, w_2, \dots, w_m$ .

$$\tilde{y} = \sum_{i=1}^m \hat{y}_i \cdot w_i \quad \text{and} \quad \sum_{i=1}^m w_i = 1.$$

In the Bayesian framework, model weights definition is straightforward. The model weights used are the posterior model weights  $w_i = p(\mathcal{M}_i|y)$  given in (S1) that represent the relative probability of each model given the data. So a major challenge is to estimate these Bayesian weights. Except for using the marginal likelihood estimations mentioned in previous section, we also investigate approximations of the weights by using the BIC for each model. In particular, model weights can be estimated through the following equations [5, 24]:

$$w_i = \frac{e^{-0.5 \cdot (BIC(\mathcal{M}_i))}}{\sum_{j=1}^m e^{-0.5 \cdot (BIC(\mathcal{M}_j))}}. \quad (\text{S11})$$

Instead of BIC, the AIC, DIC, WAIC and LooCV are also used in (S11). We investigate both approaches in insect observed rates and compare the results taking into account model complexity, data scarcity and Biological interpretation.

## 4. POWER POSTERIOR FOR GAUSSIAN AND INVERSE GAMMA DISTRIBUTION

Following [11] the power posterior in (S5) includes the likelihood raised to the power of  $t$   $p(\theta_i|y)^t$ .

In the the Gaussian case, the likelihood involved becomes:

$$p(y|\theta)^t = p^t(y|\mu, \sigma^2) = \frac{e^{-\frac{1}{2}(\sqrt{t} \cdot \frac{y-\mu}{\sigma})^2}}{\sqrt{2\pi\sigma^2}} = p\left(y|\mu, \frac{\sigma^2}{t}\right) \cdot \frac{\sqrt{t}}{t}.$$

Inserting the current ecological model  $r(T; \theta)$ , the log of the power posterior is given by:

$$\log(p(y|\theta)^t) = -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \left( \sqrt{t} \cdot \frac{y - r(T; \theta)}{\sigma} \right)^2. \quad (S1)$$

In the Inverse Gamma case, the likelihood involved becomes:

$$\begin{aligned} p(y|\theta)^t &= p(y|\alpha, \beta)^t = \frac{\beta^{\alpha \cdot t}}{\Gamma^t(\alpha)} y^{-\alpha \cdot t - t} \exp\left(-\frac{\beta \cdot t}{y}\right) = \\ &= \frac{(\beta \cdot t)^{\alpha \cdot t}}{\Gamma(\alpha \cdot t)} \frac{\Gamma(\alpha \cdot t)}{\Gamma^t(\alpha) \cdot (t)^{\alpha \cdot t}} y^{-\alpha \cdot t - 1} y^{1-t} \exp\left(-\frac{\beta \cdot t}{y}\right) = \\ &= \text{InG}(\alpha \cdot t, \beta \cdot t) \cdot \frac{\Gamma(\alpha \cdot t)}{\Gamma^t(\alpha) \cdot (t)^{\alpha \cdot t}} \cdot y^{1-t}. \end{aligned} \quad (S2)$$

Inserting the current ecological model  $r(T; \theta)$ , the log of the power posterior is given by:

$$\begin{aligned} \log(p(y|\theta)^t) &= \alpha \cdot t \cdot (\log(\alpha - 1) + \log(r(T; \theta))) - \\ &= t \cdot \log(\Gamma(\alpha)) - t \cdot (\alpha + 1) \cdot \log(y) - \frac{(\alpha - 1) \cdot (r(T; \theta)) \cdot t}{y}. \end{aligned} \quad (S3)$$

## 5. ESTIMATION OF VARIANCE OF POWER POSTERIOR METHOD

$$\begin{aligned} \hat{\sigma}_y^2 &= \text{Var}\{\log P(y|\theta)\} = \text{Var}\left\{\int_0^1 E\{\log P(y|\theta)\} dt\right\} \simeq \\ &= \sum_{i=1}^{n-1} \frac{1}{4} \text{Var}\left\{E_{\theta|y, t_i}\{\log P(y|\theta)\} + E_{\theta|y, t_{i+1}}\{\log P(y|\theta)\}\right\} (t_{i+1} - t_i)^2 \simeq \\ &= \sum_{i=1}^{n-1} \left[ \frac{sd_{\theta|y, t_i}^2}{2} \right] \cdot (t_{i+1} - t_i)^2 \end{aligned} \quad (S1)$$

where  $sd_{\theta|y, t_i}$  is the std error estimated at the  $t_i$  temperature

## 6. ESTIMATION OF VARIANCE OF IMPORTANCE SAMPLING METHOD

The marginal likelihood estimate for the  $i$ th model is given by  $m = p(y|M_i) = E_g\left\{\frac{q(\theta_i|y)}{g(\theta_i)}\right\} = \int \frac{q(\theta_i|y)}{g(\theta_i)} g(\theta_i) d\theta_i$  provided  $g(\theta_i) > 0$  whenever  $q(\theta_i|y) \neq 0$  where the density  $q(\theta_i|y)$  is equal to  $p(y|\theta_i, \phi_i) \cdot p(\theta_i, \phi_i)$  and the auxiliary importance function  $g$  used is the following:  $g(\theta_i) = g(\theta_i, \phi_i) = p(\theta_i|y) p(\phi_i|y)$

Removing the  $i$ th model index we can evaluate the marginal likelihood via MC integration which gives the formula below:

$$\hat{z}_y = \hat{p}(y) = \frac{1}{M} \sum_{j=1}^M \frac{q(\theta^j|y)}{g(\theta^j)}$$

where  $\theta^j, j = 1, 2, 3, \dots, M$  are obtained from density:  $g(\theta_i)$ .

$$E_g(\hat{z}_y) = \frac{1}{M} \sum_{j=1}^M E_g\left\{\frac{q(\theta^j|y)}{g(\theta^j)}\right\} = p(y) = m.$$

Assuming that  $\text{Cov}\left\{\frac{q(\theta^j|y)}{g(\theta^j)}, \frac{q(\theta^i|y)}{g(\theta^i)}\right\} = 0$ , for  $i \neq j$

$$V_g(\hat{z}_y) = V_g\left(\frac{1}{M} \sum_{j=1}^M \frac{q(\theta^j|y)}{g(\theta^j)}\right) = \frac{1}{M^2} \sum_{j=1}^M V_g\left\{\frac{q(\theta^j|y)}{g(\theta^j)}\right\} +$$

$$\begin{aligned} \frac{1}{M^2} \sum_{k,j=1}^M \text{Cov}_g \left\{ \frac{q(\theta^k|y)}{g(\theta^k)}, \frac{q(\theta^j|y)}{g(\theta^j)} \right\} &= \\ \frac{1}{M^2} \sum_{j=1}^M V_g \left\{ \frac{q(\theta^j|y)}{g(\theta^j)} \right\} &= \frac{1}{M} V_g \left\{ \frac{q(\theta|y)}{g(\theta)} \right\} = \frac{1}{M} \left\{ E_g \left\{ \frac{q(\theta|y)}{g(\theta)} \right\}^2 - E_g \left\{ \frac{q(\theta|y)}{g(\theta)} \right\} \right\} = \\ \frac{1}{M} \int_G \left\{ \frac{q(\theta|y)}{g(\theta)} \right\}^2 g(\theta) d\theta - \frac{1}{M} m^2 &\Rightarrow \end{aligned}$$

Also, using properties of expectation and the fact that  $g(\theta)$  is proper

$V_g(\hat{z}_y) = \frac{1}{M} \int_G \left\{ \frac{q(\theta|y) - m \cdot g(\theta)}{g(\theta)} \right\}^2 g(\theta) d\theta = \frac{\sigma_y^2}{M}$  The latest can be used to estimate the std error of the estimator  $\hat{z}_y$  via the formula:

$$\hat{\sigma}_y = \sqrt{\frac{1}{K} \sum_{j=1}^K \left\{ \frac{q(\theta^j|y)}{g(\theta^j)} - \hat{z}_y \right\}^2} \quad (S1)$$

where  $\theta^j, j = 1, 2, 3, \dots, K$  are obtained from density  $g(\theta_i)$ .

## REFERENCES

1. B. Steve, G. Andrew, J. Galin, and M. Xiao-Li, *Handbook of markov chain monte carlo* (CRC press, 2011).
2. A. Barber, C. Campbell, H. Crane, R. Lilley, and E. Tregidga, "Biocontrol of two-spotted spider mite tetranychus urticae on dwarf hops by the phytoseiid mites phytoseiulus persimilis and neoseiulus californicus," *Biocontrol Sci. Technol.* **13**, 275–284 (2003).
3. B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, "Stan: A probabilistic programming language," *J. statistical software* **76**, 1–32 (2017).
4. D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Am. Stat. Assoc.* **112**, 859–877 (2017).
5. Kass, R. E, and R. A. E, "Bayes factors," *J. american statistical association* **90**, 773–795 (1995).
6. P. J. Green, "Reversible jump markov chain monte carlo computation and bayesian model determination," *Biometrika* **82**, 711–732 (1995).
7. E. I. George and R. E. McCulloch, "Approaches for bayesian variable selection," *Stat. sinica* **7**, 339–373 (1997).
8. B. P. Carlin and S. Chib, "Bayesian model choice via markov chain monte carlo methods," *J. Royal Stat. Soc. Ser. B (Methodological)* **57**, 473–484 (1995).
9. L. Kuo and B. Mallick, "Variable selection for regression models," *Sankhyā: The Indian J. Stat. Ser. B (1960-2002)* **60**, 65–81 (1998).
10. P. Dellaportas, J. J. Forster, and I. Ntzoufras, "Bayesian variable selection using the gibbs sampler," *Biostatistics-Basel* **5**, 273–286 (2000).
11. N. Friel and A. N. Pettitt, "Marginal likelihood estimation via power posteriors," *J. Royal Stat. Soc. Ser. B (Statistical Methodol.* **70**, 589–607 (2008).
12. X.-L. Meng and W. H. Wong, "Simulating ratios of normalizing constants via a simple identity: a theoretical exploration," *Stat. Sinica* **6**, 831–860 (1996).
13. K. Perrakis, I. Ntzoufras, and E. G. Tsionas, "On the use of marginal posteriors in marginal likelihood estimation via importance sampling," *Comput. Stat. & Data Analysis* **77**, 54–69 (2014).
14. H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Autom. Control.* **19**, 716–723 (1974).
15. G. Schwarz, "Estimating the Dimension of a Model," *The Annals Stat.* **6**, 461 – 464 (1978).
16. D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde, "Bayesian measures of model complexity and fit," *J. royal statistical society: Ser. b (statistical methodology)* **64**, 583–639 (2002).

17. A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis* (CRC press, 2013).
18. S. Watanabe, "Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory," *J. Mach. Learn. Res.* **11**, 3571–3594 (2010).
19. A. Gelman and X.-L. Meng, "Simulating normalizing constants: From importance sampling to bridge sampling to path sampling," *Stat. science* **13**, 163–185 (1998).
20. A. M. Overstall and J. J. Forster, "Default bayesian model determination methods for generalised linear mixed models," *Comput. Stat. & Data Analysis* **54**, 3269–3288 (2010).
21. S. Frühwirth-Schnatter, "Estimating marginal likelihoods for mixture and markov switching models using bridge sampling techniques," *The Econom. J.* **7**, 143–167 (2004).
22. Q. F. Gronau, H. Singmann, and E.-J. Wagenmakers, "bridgesampling: An R package for estimating normalizing constants," *J. Stat. Softw.* **92**, 1–29 (2020).
23. R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2020).
24. S. T. Buckland, K. P. Burnham, and N. H. Augustin, "Model selection: an integral part of inference," *Biometrics* pp. 603–618 (1997).