Supporting Information S5: Adding observation error

Supporting Information for *How do MAR(1) models cope with hidden nonlinearities in ecological dynamics?*

Grégoire Certain, Frédéric Barraquand and Anna Gårdmark, ***Methods in Ecology and Evolution***.

Including observation error to a MAR(1) model

Formally, autoregressive models with added observation errors are referred to as *state-space* models. One strength of recent statistical packages for fitting autoregressive models in ecology, including MARSS (Holmes *et al.*, 2012; Hampton *et al.*, 2013), which we use for the analysis in the main text, is their ability to do so within a state-space framework.

Taking into account an observation error when present, which is often the case, can help to get better estimates (Lindén & Knape, 2009). However, adding an observation error, when unknown, may also bias the estimation (Knape, 2008). A recent study by Auger-Méthé *et al.* (2016) indeed demonstrated that state space models can suffer from parameter- and state-estimation problems when both the observation and process error are unknown and especially when observation errors are larger than process errors.

We therefore investigated the effect of adding observation error to the simulated datasets as follows:

$$\mathbf{y_t} = \mathbf{I}\mathbf{x_t} + \mathbf{v}_t,\, \mathbf{v}_t \sim \text{MVN}(0, \mathbf{R}) \tag{1}$$

where $\mathbf{I}$ is the identity, 2×2 matrix and the correlation structure of observation errors is specified within the matrix $\mathbf{R}$. For our purpose, $\mathbf{R}$ will be a simple scalar 2×2 matrix with diagonal elements equal to ν^2, so that process error is equal to ν^2 for the two species and errors are uncorrelated.

We limited our investigation to competitive dynamics (Gompertz, Ricker and Beverton Holt) with process error $\sigma_1^2 = 0.1$ for the sake of simplicity and feasibility (state-space models are much more time-consuming to fit than their state-only counterpart). We considered three levels of observation error, $\nu^2 = 0.01, 0.1$ and 0.5, to investigate situations in which observation error was lower than, equal to, and greater than the simulated process error, respectively.

For the fit, the observation variance-covariance matrix was specified as diagonal and equal, so that it exactly matches the way observation errors have been introduced in the simulated data. Note that in the case of process errors, a perfect match between σ_1^2 and σ_2^2, and the corresponding fitted values Σ_{11} and Σ_{22} can only be assumed when the data-generating model is of Gompertz type. When the dynamics are Ricker or Beverton-Holt, the change in nonlinearities mean that variability due to nonlinear functional forms will end up in the process error (see Fig. S5.1). That said, it is still reasonable to assume that error variances should have similar order of magnitude.

How do observation errors affect the performance of MAR(1) models?

Table S5.1 and 5.2 are structured similarly to tables 3 and 4 in the main text. Table S5.1 displays the statistics evaluating how the $\mathbf{B}$ matrix estimated in a MAR(1) model in a state-space context approaches the Jacobian matrix $\mathbf{J}$ of the underlying, data-generating model.

	model	Gompertz				Ricker				Beverton-Holt			
	Obs. Err	*None*	*0.01*	*0.1*	*0.5*	*None*	*0.01*	*0.1*	*0.5*	*None*	*0.01*	*0.1*	*0.5*
specific question	**evaluation criterion**												
Do **B** and **J** correlate ?	Corr. between b_{11} and j_{11}	0.99	0.99	0.96	0.87	0.94	0.93	0.91	0.82	0.94	0.91	0.84	0.68
	Corr. between b_{12} and j_{12}	0.87	0.79	0.57	0.30	0.92	0.58	0.52	0.21	0.91	0.54	0.45	0.33
	Corr. between b_{21} and j_{21}	0.92	0.91	0.76	0.39	0.91	0.89	0.80	0.56	0.89	0.85	0.65	0.45
	Corr. between b_{22} and j_{22}	0.99	0.98	0.95	0.80	0.94	0.82	0.77	0.52	0.89	0.81	0.67	0.37
Is there some systematic bias in the fitted **B** values ?	median of $b_{11} - j_{11}$	0.01	0.00	-0.01	0.01	0.09	0.06	0.05	0.08	0.04	0.04	0.03	0.01
	median of $b_{12} - j_{12}$	0.00	-0.01	0.02	0.09	0.07	0.05	0.08	0.17	0.00	0.00	0.04	0.11
	median of $b_{21} - j_{21}$	0.01	0.01	0.01	0.03	0.05	0.03	0.04	0.06	0.01	0.01	0.01	0.01
	median of $b_{22} - j_{22}$	-0.01	-0.01	-0.02	-0.04	0.06	0.05	0.07	0.06	-0.01	-0.01	-0.04	-0.07
Is **J** within the confidence intervals of **B** ?	% of j_{11} out of c.i.95% of b_{11}	6.33	7.33	6.67	12.00	43.33	34.33	28.00	21.33	24.00	22.67	19.33	14.67
	% of j_{12} out of c.i.95% of b_{12}	5.00	4.67	6.67	7.67	15.00	11.00	12.67	13.67	7.00	6.67	9.33	12.67
	% of j_{21} out of c.i.95% of b_{21}	6.67	7.33	8.00	9.33	29.33	19.33	20.00	20.67	10.00	9.67	9.33	11.33
	% of j_{22} out of c.i.95% of b_{22}	5.33	5.67	7.33	16.00	23.67	19.00	21.67	24.67	6.33	7.67	11.67	18.00
Do ranks in **B** values reflect ranks in **J** values ?	% of j11 correctly ranked	83.33	82.33	78.67	57.33	82.00	80.00	75.67	64.33	89.67	87.00	79.33	64.00
	% of j12 correctly ranked	77.67	74.33	67.33	45.33	77.33	73.33	67.00	49.00	88.00	88.33	74.00	51.00
	% of j21 correctly ranked	82.00	77.67	68.33	51.67	78.67	76.00	67.00	52.67	80.67	81.00	66.67	51.00
	% of j22 correctly ranked	84.33	83.33	78.33	60.33	77.33	75.00	65.00	49.33	81.67	78.33	67.00	43.67
Do signs in **B** values reflect signs in **J** values ?	% of j11 correctly signed	96.33	97.33	94.33	87.33	93.00	92.33	90.00	86.00	97.33	97.00	95.67	89.33
	% of j12 correctly signed	94.33	92.33	88.67	71.00	91.00	85.33	82.00	67.67	91.33	88.67	83.00	64.33
	% of j21 correctly signed	94.67	92.67	91.00	80.00	90.67	88.67	84.67	73.33	82.00	81.00	78.33	67.33
	% of j22 correctly signed	96.33	96.33	93.67	84.00	90.67	88.00	86.67	72.67	88.33	85.33	78.33	66.00
B and **J** matrix properties	Corr. between max. eigen. of **B** and **J**	0.93	0.90	0.81	0.60	0.64	0.62	0.53	0.39	0.90	0.83	0.74	0.70

Subjective color scale:	Best performing cases	Fairly Good	Some Issues	Less performing cases	irrelevant

Table S5.1: For each of the competition models, 4 columns are presented. The first one corresponds to the performance of MAR(1) without including observation error (as in table 3 of the main manuscript), and the three others corresponds to the three levels of observation error considered ($\nu^2 = 0.01$, 0.1 and 0.5). Rows are organized according to statistics evaluating various aspects of the approximation of $\mathbf{J}$ by $\mathbf{B}$. The colorscale is subjective but allows for quickly identifying the best performing (green) and worst performing (red) situations.

It is clear from Table S5.1 that the inclusion of observation errors makes it more challenging for

B to approximate **J**, and sometimes to a large extent. Consider for example how well b_{12} relates to j_{12}: the correlation ranges from 0.9 or 0.8 for absent or small observation errors to 0.30 for large observation errors. This points to bias in parameters (i.e., estimated parameters such as net interactions and density-dependence are away from their true values, although this bias is not the same accross simulations). The pattern is similar when considering the predictive ability of the MAR(1) model fitted with observation errors (Table S5.2).

	model	**Gompertz**				**Ricker**				**Beverton-Holt**			
	noise	***None***	***0.01***	***0.1***	***0.5***	***None***	***0.01***	***0.1***	***0.5***	***None***	***0.01***	***0.1***	***0.5***
specific question	**evaluation criterion**												
Accuracy of the detection of the environmental effect	% of $\boldsymbol{q}_{11}$ out of c.i.95% of $\boldsymbol{c}_{11}$	10.67	12.00	10.00	9.67	12.00	12.00	11.67	9.00	9.00	8.33	6.67	9.33
	% of $\boldsymbol{q}_{21}$ out of c.i.95% of $\boldsymbol{c}_{21}$	6.33	5.33	9.00	7.33	5.33	6.00	7.33	8.33	7.33	8.67	7.00	6.67
	% of $\boldsymbol{q}_{11}$ correctly signed	99.00	99.33	97.67	96.33	98.33	98.33	98.00	96.00	98.33	98.33	97.00	93.33
	median of $\boldsymbol{c}_{11}$ - $\boldsymbol{q}_{11}$	-0.01	-0.01	-0.01	0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	median of $\boldsymbol{c}_{21}$ - $\boldsymbol{q}_{21}$	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	-0.01
Accuracy of the short-term forecasts	% of well predicted next trend, sp1	84.33	85.67	84.33	84.33	82.67	82.33	81.67	80.33	74.67	74.33	73.33	72.67
	% of well predicted next trend, sp2	78.00	78.00	79.33	78.00	77.33	75.33	74.67	74.00	69.00	69.33	68.67	63.33
	% ASE >1, sp1	9.33	9.00	11.00	14.33	18.00	20.00	22.00	27.67	23.67	25.67	27.00	33.67
	%ASE >1, sp2	17.00	17.00	18.00	25.33	28.00	29.00	29.67	31.00	31.00	31.00	31.33	35.33
Accuracy of the prediction of the PRESS effect	Correlation between Δx_1^* and Δn_1^*	0.98	0.97	0.96	0.92	0.48	0.57	0.55	0.58	0.70	0.71	0.72	0.74
	Correlation between Δx_2^* and Δn_2^*	0.93	0.92	0.91	0.85	0.72	0.70	0.62	0.51	0.59	0.57	0.52	0.40
	Slope of the linear regression $\Delta x_1^* \sim \Delta n_1^*$	0.92	0.92	0.88	0.76	6.63	5.88	5.18	3.73	4.46	4.32	3.83	3.16
	Slope of the linear regression $\Delta x_2^* \sim \Delta n_2^*$	0.89	0.89	0.92	0.72	0.45	0.45	0.43	0.39	0.32	0.31	0.31	0.25
	% of Δn_1^* out of c.i.95% of Δx_1^*	3.68	2.34	3.67	2.33	49.50	42.48	30.33	11.07	37.00	33.33	22.33	8.33
	% of Δn_2^* out of c.i.95% of Δx_2^*	2.34	1.67	2.00	3.33	10.37	6.69	4.00	6.04	22.00	16.33	13.00	4.67
	% of Δn_1^* correctly signed	99.00	99.00	98.33	97.33	97.66	96.99	97.67	96.31	98.33	98.33	95.67	95.33
	% of Δn_2^* correctly signed	88.29	88.96	81.67	74.67	86.29	83.61	78.33	66.78	78.67	77.67	74.00	69.33

Subjective color scale:	Best performing cases	Fairly Good	Some Issues	Less performing cases	irrelevant

Table S5.2: For each of the competition models, 4 columns are presented. The first one corresponds to the performance of MAR(1) without including observation errors (as in table 4 of the main manuscript), and the three others corresponds to the three levels of observation errors considered ($\nu^2 = 0.01$, 0.1 and 0.5). Rows are organized according to statistics evaluating the performance of the short-and long-term predictions offerd by MAR(1)**.** The colorscale is subjective but allows for quickly identifying the best performing (green) and worst performing (red) situations.

Neither notable improvements nor strong degradations in short-term forecasts or long-term predictions are associated with the inclusion of observation errors, thus observation error identifiability issues mostly impact parameter inference.

How well are observation errors recovered by MAR(1) models?

We investigated how well the state-space formulation of MAR(1) models would distinguish observation error from process error by comparing estimated values for the fitted observation error ν^2 and fitted process errors Σ_{11} and Σ_{22} among simulations (Fig S4.1).

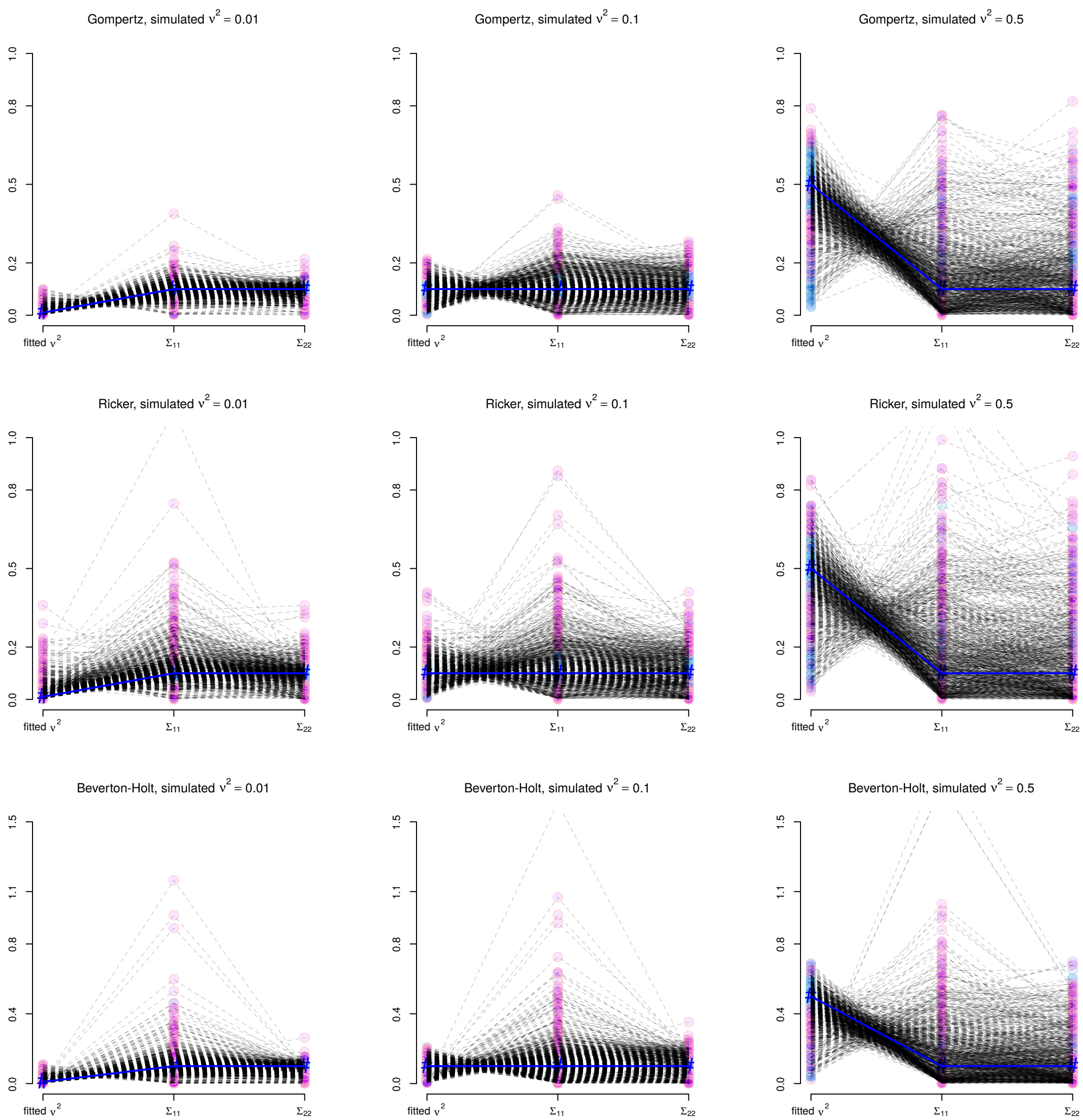

Figure S5.1: This figure shows, for the three competition models investigated (rows, Gompertz, Ricker and Beverton Holt) and the three observation error values simulated (columns, $\nu^2 = 0.01$, 0.1 and 0.5) how well those are approximated by fitted observation errors and fitted process errors Σ_{11} and Σ_{22} for 1000 simulated experiments. Note that Σ_{11} and Σ_{22} are the fitted counterparts of process errors σ_1^2 and σ_2^2, included in species 1 and 2 growth rates. For this experiment focusing on observation errors, both were kept at 0.1. Each dashed line in the graph corresponds to a single simulation with a unique parameter set. Red dots correspond to cases in wich the true simulated value is outside of the confidence intervals provided by the MAR(1) model.

Figure S5.1 shows that even for data generated by Gompertz dynamics, the estimated process and observation errors tend to trade-off: whenever the observation error is underestimated the process error is overestimated and vice versa, especially for species 1 that is also under the influence of environmental forcing. These findings confirm results obtained in other modelling studies (Knape, 2008; Auger-Méthé *et al.*, 2016). Even though not accounting for observation errors when present can possibly lead to biased estimates of the environmental effect (Lindén & Knape, 2009), doing so without prior knowledge of the magnitude of observation error versus process error to constrain model fitting is a risky endeavour: observation errors might end up in the process error term, or vice versa, which may significantly hamper conclusions drawn from fitted variances or estimated model coefficients.

References

Auger-Méthé, M., Field, C., Albertsen, C., Derocher, A., Lewis, M., Jonsen, I. & J., M.F. (2016) State-space model's dirty little secrets: even simple linear gaussian models can have estimation problems. *Scientific Reports*, **6**, 1–10.

Hampton, S., Holmes, E., Scheef, L., Scheuerell, M., Katz, S., Pendleton, D. & Ward, E. (2013) Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. *Ecology*, **94**, 2663–2664.

Holmes, E.E., Ward, E.J. & Wills, K. (2012) MARSS: Multivariate Autoregressive State-space Models for Analyzing Time-series Data. *The R Journal*, **4**, 11–19.

Knape, J. (2008) Estimability of density dependence in models of time series data. *Ecology*, **89**, 2994–3000.

Lindén, A. & Knape, J. (2009) Estimating environmental effects on population dynamics: consequences of observation error. *Oikos*, **118**, 675–680.