This is an electronic appendix to the paper by Stenseth *et al*. 2004 Modelling non-additive and non-linear signals from climatic noise in ecological time series: Soay sheep as an example. *Proc. R. Soc. Lond.* B **271**, 1985–1993. (doi:10.1098/rspb.2004.2794)

Electronic Appendix A

Generating synthetic time series from the Age-Structured Markov Model (ASMM)

Following the previous work of Coulson *et al.* (2001), an age-structured Markov model (ASMM) was used to generate time series under different climatic regimes. The summer population size at time *i* was predicted from the previous summer population combining the processes of survival and fertility for each sex/age-class combination. Survival and fertility probabilities were estimated from recaptures and recoveries of marked individuals from 1986 to 2000. Collection of extra data since the construction of the ASMM published by Coulson *et al.* (2001) led to refine this ASMM. The difference with the model of Coulson *et al.* 2001) is in the way fecundity has been modelled (see below). We also considered a full age structure (20 and 10 classes for females and males (Coulson *et al.* 2001), respectively (see Table 3, 4). The North Atlantic Oscillation Index (NAO) was used as the only density-independent predictor of survival and fertility considered, which contrasts with Coulson *et al.* (2001) who considered more climatic variables and included February and March rainfall as predictors of female yearling and adult survival in the most parsimonious model. Measures of March and February rainfall were available from 1956 onwards only and we were not able to use them for our purposes. We first simulated 50 time series of 50 years using real consecutive NAO values. Each series started from a randomly selected year between 1864 and 1951. Then, in a larger-scale simulation study reported in Table 2, we simulated 1,000 time series of 50 years with the (centered) NAO drawn from a normal distribution with zero mean and standard deviation equal to c times 0.2, the observed standard deviation, where c equals a number of chosen values; see Table 2. Previous summer population size was also used as an external covariate to account for density-dependent effect as in Catchpole *et al.* (2000) and Coulson *et al.* (2001). Covariates were centred as follows:

Population density = (actual population density – 1202.86)/100
NAO= (actual NAO – 1.73)/10

The population size in the summer *i*, was the sum of adults of age *j* present at *i*-1 and estimated to be alive at *i*, and the total number of lambs produced and alive in summer *i*. We introduced demographic stochasticity by considering the number of lambs produced at *i* and the individuals alive at year *i* a random number chosen from a binomial distribution with parameters ($N_{j,i-1}$, $\theta_{j,i-1}$) where $N_{j,i-1}$,is the number of individuals, or of mothers in the case of fertility probabilities, of age *j* present at *i*-1 and θ is the parameter of survival or fertility during the interval *i*-1, *i*.

SURVIVAL PROBABILITY

Survival probability was estimated by maximum likelihood procedure from the simultaneous analysis of recaptures and recoveries of marked animals. This analysis uses data up to 2001; full details of the procedure are given in Catchpole *et al.* (2000). We first grouped those age classes that shared similar parameters (Catchpole *et al.* 2000). For each age group, the survival, ϕ, was modelled as a function of the North Atlantic Oscillation index (NAO), the previous summer population size (*population size*) and their statistical interaction (NAO·*population size*) as:

$$\text{logit}(\phi)=\beta_0+\beta_1\,(\textit{population size})+\beta_2\,(\text{NAO})+\beta_3(\text{NAO}\cdot\textit{population size})$$

where β_x is the linear predictor of the effect considered. Only significant effects were retained (Table 3).

Tab.3 Linear predictors of survival probability according to age classes

Age-class	Age	Intercept	NAO	POP	NAO·POP
Females					
1	1	0.5403	-1.6086	-0.3078	-0.6602
2	2	2.2797	-2.4922	-0.1924	-0.5816
3	3-7	2.7725	-1.9750	-0.1702	-0.5041
4	>7	1.6199	-1.2312	-0.2409	-1.3160
Males					
1	1	-0.2068	-3.5837	-0.3053	-0.4202
2	2-7	3.4038	-14.7928	-0.5066	1.6893
3	>7	-0.4812	-	-	-

FERTILITY

Births generally occurred in April when each female could give birth to up to two lambs. Female Soay sheep can give birth at their first birthday but they never produce twins. Fertility, i.e. the number of lambs that survive until the summer per female, was modelled as the product of three distinct probabilities: a) the probability of giving birth, b) the probability of producing a single lamb and c) the neonatal survival (the probability that the lamb surviving the first three months of life). Similarly to that described above, for each age group, each probability was modelled as a function of three covariates as:

$$\text{logit}(\theta)=\beta_0+\beta_1\,(\textit{population size})+\beta_2\,(\text{NAO})+\beta_3(\text{NAO}\cdot\textit{population size})$$

where θ is the parameter of interest and βs are the linear predictors. Only significant effects were retained (Table 4). The probability of producing a single lamb was considered constant through time but varied as function of the age of the mother (Table 5).

Tab.4 Linear predictors of giving birth and neonatal survival probability from the retained models

Age-class	Age	Intercept	NAO	POP	POP·NAO
Giving birth probability					
1	1	-0.915	-2.069	-0.376	-
2	2	0.815	-2.085	-0.1017	-
3	3_7	1.3869	-	-0.0797	-
4	8_10	1.106	-2.052	-1.09	-0.812
5	>10	-1.099	-	-	-
Neonatal survival					
1	1	-0.654	-2.313	-0.3436	-
2	2	1.293	-3.55	-0.2318	-
3	3_10	2.084	-1.433	-0.0614	-0.562
4	>10	0.887	-	-	-

Tab.5 Probability of producing single lambs

Age of the mother	Probability of producing a single lamb
1	1
2	1
3	0.94
4	0.89
5	0.83
6	0.77
7	0.74
8	0.73
9	0.75
10	0.80
11	0.86
12	1

Using R^2 and AIC for model comparison

The adjusted R^2 and the AIC are two popular model selection criteria, but they involve different sets of assumptions and measure different aspects of a model under consideration. The adjusted R^2 is defined as 1-(n-1)RSS/[(n-p)SSE] where n is the sample size, p is the number of independently adjusted parameters; RSS is the residual sum of squared errors and SSE is the sum of squared deviations of the data from the mean. Under very general assumptions including homogeneous noise variance, the adjusted R^2 is a consistent estimator of the percent of variance due to the (true) 1-step predictor of the model whereas the AIC is an asymptotically unbiased estimator of the entropy (a dissimilarity measure) of the true distribution of the data relative to that according to the model. In general, the derivation of the AIC requires a number of strong assumptions such as asymptotic normality of the maximum likelihood estimator (Tong 1990). While the AIC properly accounts for heteroscedastic noise variance, it must be interpreted cautiously when the number of data in any regime is small in which case the normality assumption of the estimator may not obtain. In contrast, the adjusted R^2 is a global measure of the prediction accuracy of the model that ordinarily requires less restrictive assumptions, see Eubank (1999) for a related discussion of linear predictors. Nevertheless, the AIC enjoys an interesting “consistency” property. Suppose that the true model has an infinite number of parameters [e.g., if the true B function is a (natural cubic) spline function with infinite d.f. (knots)], and suppose that we select the best model out of all models with B as a spline function that has d.f. ranging from 1 to n^{α}, where n is the sample size and $\alpha<1$ describes the rate of complexity of model that is entertained with increasing amount of data. Then, by analogy of a “simpler” situation (c.f. section 2.3 of Eubank 1999), it is expected that the d.f. selected by the AIC converges to the optimal d.f. minimizing the expected square prediction errors; furthermore, the convergence rate is of the order $n^{(\alpha-1)/2}$. This “consistency” property of AIC is particularly relevant as the form of the B function is generally unknown. On the other hand, in the idealistic case that the true B function is a spline with finite number of d.f., the AIC is known to overfit the model with a finite probability; when the class of models include the true model and the number of redundant parameters is large, the probability of overfitting approaches 0.29 for large sample size. Thus, if the true B function is of d.f. = 1 (i.e., the NAO does not affect the process), the AIC will have probability close to 0.29 of selecting d.f. greater than 1 if a large range of d.f. is entertained in the model selection and if sample size is large. The use of AIC in detecting the NAO effect may therefore be criticized owing to this high false positive rate. However, this criticism is only valid if we are testing the NAO effect against a scientific gold standard in the form of the model with d.f. = 1. In the case that such a gold standard is absent and when the true model is unlikely to be within the class of models under study, the consistency property of the AIC seems appealing enough to warrant its use in exploring possible effect of the NAO on an underlying system; furthermore, the AIC (Burnham *et al.* 1995) tends to select models with small biases and variance in the parameter estimates, as well as good balance between errors of under- and over-fitting. As with all time-series analysis, our statistical approach is only able to establish an association between the NAO and the process suggestively through the B function, but experiments and other means are necessary to more rigorously establish the mechanistic link.

The AIC (Table 1) indicates that $B(NAO_t)$ may be a linear function of the NAO; the corresponding constrained TAR model has AIC = -125.5. On the other hand, the criterion of the adjusted R^2 suggests that B may be a non-linear function approximated by a natural spline function with 4 d.f.; indeed, the adjusted R^2 equals 13.2% and 12.0% when the d.f. of the spline function fit for B is increased to 5 and 6 (unreported), respectively. (The corresponding AIC are -122.5 and -120.9.) However, the shape of the fitted B-function (unreported) is almost identical when the d.f. is between 4 and 6. The divergence of the conclusions from the two criteria partly owes to the fact that the adjusted R^2 measures the global prediction accuracy whereas the AIC takes into account that the noise variances differ in the two regimes (*i.e.,* whether or not the population size in the preceding year exceeds the threshold k). The constrained FCTAR model with b being a constant has AIC = -124.7, R^2 = 10.3% and adjusted R^2 = 6.0%, whereas the unconstrained SETAR model of Grenfell *et al.* (1998) has AIC = -125.0, R^2 = 17.5% and adjusted R^2 = 11.5%, with both models fitted over the period of 1955-2000. It may properly be argued (p. 248 in Tong 1990 and see below) that we should count the threshold parameter of a discontinuous SETAR model as 2 parameters rather than simply counting it as one parameter as it is done in the above computation of AIC and R^2. Our argument is based on a likelihood analysis of a related testing problem where the general model is $x_t = a_0 + a_1 I(x_{t-1} \leq k) + \varepsilon_t$ where the noise term ε is normally distributed; $I(\cdot)$ equals 1 if the enclosed expression is true and 0 otherwise. The null hypothesis specifies that $a_1 = 0$. If k is unknown, the tail of the distribution of the likelihood ratio test is approximated by that of χ^2

with 3 d.f. under the null hypothesis of constant regression function, for large sample size, whereas if k is known, the likelihood ratio test is approximately χ^2 with 1d.f. This suggests that the threshold parameter k is worth 2 parameters.

With the threshold parameter counted as two parameters, the SETAR model of Grenfell *et al.* (1998) has AIC = -123.0 and adjusted R^2 = 9.2%. If we use the adjusted R^2 as the model selection criterion, our constrained FCTAR model with b = B(NAO) modelled by natural splines with 4 or 5 d.f. outperforms the unconstrained SETAR model of Grenfell *et al.* 1998. Although in terms of either model selection criterion, the unconstrained SETAR model of Grenfell *et al.* (1998) is highly competitive, constraining the TAR model as suggested above (a_1 = 1) and incorporating b = B(NAO), our constrained FCTAR model is indeed the better one; obviously the critical point here is the inclusion of b = B(NAO) – the core of this paper. Finally, notice that the mechanistic model developed by Coulson *et al.* (2001) has over 90% R^2; however, due to the high parameter-per-data ratio, the high R^2 may be somewhat biased; unfortunately an unbiased estimate of the predictive performance of this mechanistic model will only be available after more – and future – data are collected.

References

Burnham, K. P., White, G. C. & Anderson, D. R. 1995 Model selection strategy in the analysis of capture-recapture data. *Biometrics* **51**, 888-898.

Catchpole, E. A., Morgan, B. J. T., Coulson, T. N., Freeman, S. N. & Albon, S. D. 2000 Factors influencing Soay sheep survival. *Appl Stat* **49**, 453-472.

Coulson, T., Catchpole, E. A., Albon, S. D., Morgan, B. J. T., Pemberton, J. M., Clutton-Brock, T. H., Crawley, M. J. & Grenfell, B. T. 2001 Age, sex, density, winter weather, and population crashes in Soay sheep. *Science* **292**, 1528-1531.

Eubank, R. L. 1999 *Nonparametric Regression and Spline Smoothing.* New York: Marcel Dekker.

Grenfell, B. T., Wilson, K., Finkenstädt, B. F., Coulson, T. N., Murray, S., Albon, S. D., Pemberton, J. M., Clutton-Brock, T. H. & Crawley, M. J. 1998 Noise and determinism in synchronized sheep dynamics. *Nature* **394**, 674-677.

Tong, H. 1990 *Non-linear time series: a dynamical system approach.* Oxford: Clarendon Press.