English Translation of "Dependance de Facteurs de Milieu dans les Estimations

de Taux de Survie par Capture-Recapture"

by J. Clobert and J. D. Lebreton,

Biometrics, 41, 1031-1037

Translation by Dominique Latour, North Carolina State University

Estimation of the relationship between environmental factors and survival rates in capture-recapture experiments.

1. Introduction

The general capture and recapture models often lack parsimony, which results in poor precision of parameter estimates; however, these models often lead to explicit estimators of survival rate. This is the case in particular for models based on recaptures of live animals (Cormack, 1964) or on recoveries of dead animals (Seber, 1970, 1971). Models with fewer parameters, based on hypotheses of constancy through time, exist for both approaches. Tests permitting us to choose the most adequate model for a given data set are also available [Sandland and Kirkwood (1981) and Clobert et. al. (1985) for recaptures, Seber (1973) and Lebreton (1977) for recoveries of dead animals].

However, when the survival rate is strongly dependent on the variations of an environmental factor, models for the analysis of such a relationship have only been developed for the case of tagged animals found dead (North and Morgan, 1979); they proposed a model which expresses the survival rate as a logistic function of an external variable. The purpose of this article is to present a similar approach for recaptures of live animals, regardless of the type of recapture.

2. Multinomial model for recapture histories.

Let s_i be the probability that an individual survives from year i-1 to i and p_i the probability that an individual alive in year i is captured in the ith year. Under the hypotheses of independence between individuals and absence of emigration, the numbers in the different capture-recapture histories for individuals of the same cohort (there are 2^j capture-recapture histories for j years of observation) have a multinomial distribution. The likelihood for n consecutive cohorts (leading respectively to j = n, n-1,...,1,

years of recapture) can be written as a product of multinomial densities. Apart from constant terms the likelihood can be reduced to [see Cormack (1964) for more details]:

$$L = \prod_{i=1}^{n} s_i^{v_i} p_i^{d_i} \left(1 - p_i\right)^{v_i - d_i} \left[\sum_{k=i-1}^{n} \left\{\prod_{j=i}^{k} \left[\left(1 - p_j\right) s_j\right]\right\} \left(1 - s_{k+1}\right)\right]^{c_i}$$

with

$$\prod_{j=1}^{k} \left(1-p_j\right) s_j = 1 \qquad \text{for } i > k$$

and $s_{n+1} = 0$ where

$\quad d_i$ is the number of distinct individuals captured in the $i+1^{th}$ year.

$\quad v_i$ is the number of individuals known to be alive after year i,

$\quad c_i$ is the number of individuals seen for the last time in year i. This general model due to Cormack (1964), and denoted $\left(p_t, s_t\right)$ by Sandland and Kirkwood (1981), allows us to obtain explicity the maximum likelihood estimates of $s_1, \ldots, s_{n-1}, p_1, \ldots, p_{n-1}$, and of the product $s_n p_n$ (Cormack, (1964). Several other models have been derived recently under constraints of equality of survival rates $\left(\text{model } p_t, s\right)$, of sighting rates $\left(\text{model } p, s_t\right)$ or of both (model p,s) depending on whether the probabilities of capture or of survival are constant or variable with respect to time (Sandland and Kirkwood, 1981; Clobert et. al., 1985).

3. Models where survival is dependent on external variables

In order to model the relationship between the survival rate and q external variables, we consider several models in which the survival rate between i-1 and i is expressed either as a linear function $s_i = a_0 + \sum_{j=1}^{q} a_j x_j(i)$ of the external variables, or as a logistic-linear

3

function

$$s_i = \exp\left[a_0 + \sum_{j=1}^{q} a_j x_j(i)\right] \Big/ \left\{1 + \exp\left[a_0 + \sum_{j=1}^{q} a_j x_j(i)\right]\right\},$$

denoted for convenience by $s_i = \ell\left[a_0 + \sum_{j=1}^{q} a_j x_j(i)\right]$.

The sighting rate may depend on time - models

$\left[p_t, \ a_0 + \sum_{j=1}^{q} a_j x_j(t)\right]$ and $\left[p_t, \ \ell\left[a_0 + \sum_{j=1}^{q} a_j x_j(t)\right]\right]$ respectively - or

be constant - models $\left[p, \ a_0 + \sum a_j x_j(t)\right]$ and $\left[p, \ell\left[a_0 + \sum a_j x_j(t)\right]\right]$

respectively.

The logistic relationship has the advantage that it always predicts survivals lying between 0 and 1, while being fairly close to the linear relationship when the probability is between 0.2 and 0.8 (see Cox, 1972, Table 2.1; and North 1979, Figure 1, p. 25).

The derivatives of the log-likelihood function $\text{Ln}L$ with respect to a_0 and a_j ($j \geq 1$) are obtained from the derivatives with respect to s_i and p_i in the model $\left[p_t, \ s_t\right]$ (given in Clobert et. al., 1985):

in the linear case:

$$\frac{\partial \text{Ln}L}{\partial a_0} = \sum_i \frac{\partial \text{Ln}L}{\partial s_i} \quad \text{and} \quad \frac{\partial \text{Ln}L}{\partial a_j} = \sum_i \frac{\partial \text{Ln}L}{\partial s_i} x_j(i)$$

in the logistic case:

$$\frac{\partial \text{Ln}L}{\partial a_0} = \sum_i \frac{\partial \text{Ln}L}{\partial s_i} s_i(1-s_i) \quad \text{and} \quad \frac{\partial \text{Ln}L}{\partial a_j} = \sum_i \frac{\partial \text{Ln}L}{\partial s_i} x_j(i) s_i(1-s_i),$$

where s_i is replaced respectively by $a_0 + \sum_j a_j x_j(i)$ and $\ell\left[a_0 + \sum_j a_j x_j(i)\right]$.

Since they cannot be found explicity, the maximum likelihood estimators of the a_j's are obtained numerically (subprogram VA09AD of the Harwell Library, Fletcher's Method, 1970). The matrix of descent, updated at

each iteration, converges to the matrix of second derivatives at the point $\left(\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_q, \hat{p} \text{ or } \hat{p}_t\right)$. Following the usual theory of maximum likelihood, its inverse provides an estimate for the asymptotic variances and covariances of the parameters.

Likelihood ratio tests have been chosen to compare the different models. For example, in the case where the survival is linearly dependent on a single external variable:

$Y_1 = 2\text{Ln}L\left(p_t \text{ or } p, a_0 + a_1 x(t)\right) - 2\text{Ln}L\left(p_t \text{ or } p, s\right)$ provides a test for the null hypothesis $a_1 = 0$ (Y_1 is asymptotically distributed as a χ^2 with 1 degree of freedom).

$Y_2 = 2 \text{Ln}L\left(p_t \text{ or } p, s_t\right) - 2 \text{Ln}L\left(p_t \text{ or } p, a_0 + a_1 x(t)\right)$ provides a test for the linearity of the relationship (Y_2 is asymptotically distributed as a χ^2 with n-2 degrees of freedom in the case of equality of the sighting rates, n-3 in the other case).

$Y_3 = 2 \text{Ln}L\left(p_t, a_0 + a_1 x(t)\right) - 2 \text{Ln}L\left(p, a_0 + a_1 x(t)\right)$ permits us to test that capture rates are constant provided the linear relationship of the survivals is accepted (Y_3 is asymptotically distributed as a χ^2 with n-1 degrees of freedom) and, because of greater parsimony, with probably greater power than for the analogous test between the models $\left(p_t, s_t\right)$ and $\left(p, s_t\right)$. In other cases (logistic relationship, many variables), the likelihood ratio tests can be easily derived, note, however, (cf. the statistic Y_2 from above) that in the model $\left(p_t, s_t\right)$ there are only 2n-1 estimable parameters.

4. Example

In the context of a study of the dynamics of a population of Starling,

5

Sturnus vulgaris, 360 nests were followed from 1976 to 1982 in a wooded

area in Belgium. Only the females which nested were tagged and monitored.

There are only n=6 years of recapture since 1976 was the first year of

tagging. For the Starling, the survival rate becomes stable at one year

of age, in other words at the first reproduction, which justifies in part

the hypothesis that all the individuals studied have the same survival,

whatever their age.

Moreover, the Starling is almost always faithful to the previous nesting

site: 20% return to the same nest, 100% to within 700 meters (Clobert, Ph.D.

thesis, Université Catholique de Louvain, Louvain – la– Neuve, Belgium,

1981). In order to limit the disturbance on nesting, all the nests were

visited only once a season; this justifies the hypothesis that the recapture

rates are the same for all individuals.

A first analysis of the data (Table 1, col. 1-5) suggested (Clobert

et. al., 1985) selection of the model (p, s_t), that is to say a year-

specific survival rate and a constant probability of capture (Table 1, col.

7-11).

The winter of 1978-1979 (denoted 79 on figure 1), one of the coldest in

50 years in Belgium, corresponds to the lowest survival rate in the model

(p, s_t) ($\ddot{s}_3 = 0.37$; Table 1); it was then logical to believe that the

weather had an effect on the survival rates, as is often the case with species

having generations of short duration. The adults of the population studied

being winter resident mainly in Basse and Middle Belgium and Northern France

(Clobert, Ph.D. Thesis, 1981; Roggeman, pers. comm.), we used meteorologic

statistics of the Brussels area (Uccle Station), one hundred km. west of the

zone studied. In order to cover as much as possible of the non-breeding

season, we used the period from October (end of the partial migration) to

March (beginning of returns). Finally, in order to look at major climatic trends, we used the average temperature. In summary, it is the mean temperature from October to March ($^{\circ}$C) in the region of winter residence (Table 1, Col. 5) which is used as the external variable $x(i)$. We have carefully selected this variable a priori and have not analyzed any other variables (average minimum temperature, other periods) which would have probably produced similar results given the small number of points and the generally high correlations between climatic variables.

The model $(p, a_0 + a_1 x(t))$ yields then the estimators (figure 1):

$$\tilde{s}_i = -0.189 + 0.133 \ x(i) \text{ and } \tilde{p} = 0.747.$$

The matrix of estimated asymptotic variance–covariances for $(\tilde{a}_0, \tilde{a}_1, \tilde{p})$ is:

$$
\begin{pmatrix}
0.014190 & \text{------} & \text{------} \\
-0.002681 & 0.000518 & \text{------} \\
0.000557 & -0.000147 & 0.000849
\end{pmatrix}
$$

The estimates of the asymptotic correlations between parameters are:

$$r_{\tilde{a}_0, \tilde{a}_1} = 0.989, \ r_{\tilde{a}_0 \tilde{p}} = +0.160, \text{ and } r_{\tilde{a}_1 \tilde{p}} = -0.222.$$

In spite of the high correlation between $\tilde{a}_0$ and $\tilde{a}_1$, we did not encounter problems of convergence, yet reparameterization might obviously be useful in certain cases (the correlation between $\tilde{a}_1$ and $\tilde{a}_0 + \tilde{a}_1 \bar{x}$ is only 0.296).

Table 1

Recapture data for Starlings (sturnus vulgaris)

Year	i	b_i	d_i	c_i	x_i	$\hat{s}_i$	$\hat{p}_i$	$\hat{\sigma}(\hat{s}_i)$	$\ddot{s}_i$	$\hat{\sigma}(\ddot{s}_i)$
1977	1	52	27	21	5.98	0.66	0.79	.079	0.67	.079
1978	2	204	122	102	6.20	0.65	0.78	.049	0.67	.043
1979	3	102	74	151	4.05	0.37	0.79	.037	0.37	.035
1980	4	126	95	101	5.97	0.59	0.73	.059	0.58	.050
1981	5	73	67	102	5.27	0.53	0.62	.053	0.48	.044
1982	6	43	48	75	4.97	——	——	——	0.46	.058

b_i number of individuals marked in year i (for d_i and c_i, see the text).

x_i average winter temperature from October to March in ^oC.

$\hat{s}_i$, $\hat{p}_i$ estimates of survival and sighting rates for the model (p_t, s_t).

$\ddot{s}_i$ survival rate estimate for the model (p, s_t) $\left(\ddot{p} = 0.743 \pm 0.029\right)$.

$\hat{\sigma}$, $\ddot{\sigma}$ standard deviation respectively of $\hat{s}_i$ and $\ddot{s}_i$.

Table 2

Likelihood ratio tests between different recapture models.

Models	Number of Parameters	$-2\mathrm{Ln}L$	Null hypothesis	Test	Conclusion
p_t, s_t	11	1625.6			
p, s_t	7	1629.6	$p_t = p \; \forall t$	$Y_0 = 1629.6$ -1625.6 $=4.0$ $< \chi^2_{.95}(4)$	$p_t, s \gtrless p_t, s_t$ hypothesis $p_t = p$ accepted
$p, a_0 + a_1 x(t)$	3	1633.0	$p_t = p,$ $s_t = a_0 + a_1 x(t)$	$Y_2 = 1633.0$ -1629.6 $=3.4$ $< \chi^2_{.95}(4)$	$p, a_0 + a_1 x(t) \gtrless p, s_t$ hypothesis $s_t = a_0 + a_1 x(t)$ accepted
$p_t, a_0 + a_1 x(t)$	8	1626.1	$s_t = a_0 + a_1 x(t) \forall t$	$Y_3 = 1633.0$ -1626.1 $=6.9$ $< \chi^2_{.95}(5)$	$p, a_0 + a_1 x(t)$ $\gtrless p_t, a_0 + a_1 x(t)$ hypothesis $p_t = p$ accepted
p, s	2	1665.8	$s_t = s, p_t = p \; \forall t$	$Y_1 = 1665.8$ -1633.0 $=32.8$ $> \chi^2_{.99}(1)$	$p, s \gtrless p, a_0 + a_1 x(t)$ hypothesis $s_t = s$ rejected

The tests used are partially redundant and the test using Y_3 is an advantageous substitute for the test based on Y_0; once the hypothesis $s_t = a_0 + a_1 x(t)$ is accepted.

The null hypothesis $a_1=0$ (Table 2) is rejected:

$$Y_1 = 1665.8 \left[\text{model}(p,s)\right] - 1633.0 \left[\text{model }(p,a_0+a_1x(t))\right] = 32.8$$

$> \chi^2_{0.99}(1) = 6.63$. The linearity of the relationship is not rejected:

$$Y_2 = 1633.0 \left[\text{model}\left(p,a_0+a_1x(t)\right)\right] - 1629.6\left[\text{model}\left(p,s_t\right)\right] = 3.4 < \chi^2_{0.95}(4) = 9.49.$$

Finally, the hypothesis of constancy of capture probabilities is also

accepted: $Y_3 = 1633.0\left[\text{model}\left(p,a_0+a_1x(t)\right)\right] - 1626.1\left[\text{model}\left(p_t,a_0+a_1x(t)\right)\right]$

$$= 6.9 < \chi^2_{0.95}(5) = 11.07 .$$

The logistic-linear model yields results almost identical to those from

the linear model (to such an extent that we did not plot both curves on

figure 1), as shown by the similarity of the predicted survivals (the

largest difference being 0.0002) and of the likelihood values $\left[-2\text{Ln}L=1633.028\right.$

for the model $\left(p,\ell\left(a_0+a_1x(t)\right)\right.$ against 1633.018 for the model $\left.\left(p,a_0+a_1x(t)\right)\right]$.

If we had to extrapolate, the logistic-linear model

$\left(\text{with } \hat{a}_0 = -2.833 \pm 0.5198 \text{ and } \hat{a}_1 = 0.5478 \pm 0.0987\right)$ would obviously be

preferable, although with the linear model, the prediction of a negative

survival or survival larger than 1 would require a mean temperature $x(i)$

lower than 1.42°C and higher than 8.64°C respectively, values which are very

unlikely since the extreme values observed since 1961 were 3.46°C and 6.35°C.

The shape of figure 1 might suggest a quadratic relationship: the

introduction of the square of the temperature improves the fit only slightly

whether in the linear model $\left[\Delta\left(-2\text{Ln}L\right) = 1.929 < \chi^2_{0.95}(1) = 3.84\right]$ or in the

logistic-linear $\left[\Delta\left(-2\text{Ln}L\right) = 1.877\right]$ but causes an important loss of precision

(see figure 1).

The estimates $\tilde{s}_i$ and $\tilde{p}$, obtained from fitting the model $\left(p,a_0+a_1x(t)\right)$,

allow computation of the expected sizes in the $\sum\limits_{j=1}^{n} 2^j = 2^{n+1}-2 = 126$

elementary cells, which have been grouped according to the date of the last
capture within each cohort. In this way, we are grouping individuals most
likely to die at the same time, since not observed after this date, which
tests the quality of the $\tilde{s}_i$, while keeping the emphasis on possible
differences of behavior between cohorts. The model $\left(p, a_0 + a_1 x(t)\right)$ is accepted
$\left[\chi^2_{obs} = 13.13 < \chi^2_{0.98}(18) = 25.99\right]$. Thus, survival rate directly
proportional to the mean temperature from October to March and a constant
capture rate adequately describe the data.

5. Conclusion

In this example, the proposed model presents two important advantages.
First, it allows an analysis in good statistical conditions of the
relationship between the survival rate and an external variable. Indeed,
the direct analysis using simple linear regression between the estimates
of the survival rates and the variable is biased by the autocorrelation
between successive estimates of the survival rate inherent to time—specific
capture—recapture models. The likelihood ratio test between the models
$\left(p \text{ or } p_t, s_t\right)$ and $\left(p \text{ or } p_t, a_0 + a_1 s(t)\right)$ replaces advantageously the test
that the slope of the regression line is zero.

Secondly, it permits significant reduction of the number of parameters
to be estimated (greater parsimony and hence better accuracy). In the
above example, for 600 individuals tagged, and 6 years of observation hence
11 identifiable parameters, 3 parameters suffice to describe adequately
the set of observations (nonsignificant goodness—of—fit—test).

Finally, if indicated by the data, our models allow use of more than one
variable, in an analogous fashion to multiple regression, in spite of the
unavoidable problems associated with selecting the variables and with

11

validating the results. A FORTRAN 77 program which allows fitting of the various stated models is available from the authors.

Acknowledgments

We thank two anonymous readers for their constructive comments.

REFERENCES

Clobert, J. (1981). *Étude de la dynamique d'une population d'Étourneaux sansonnets* (Sturnus vulgaris, L.). Thèse de doctorat, Université Catholique de Louvain, Louvain-la-Neuve, Belgique.

Clobert, J. (1982). *Étude des taux de survie chez l'Étourneau, calculés à partir de données de capture-marquage-recapture. Le Gerfaut* 72, 255-285.

Clobert, J., Lebreton, J. D., Clobert-Gillet, M., et Coquillart, H. (1985). The estimation of survival in bird populations by recaptures or sightings of marked individuals. In *Statistics in Ornithology*, B. J. T. Morgan and P. M. North (eds), 197-213. New York: Springer Verlag.

Cormack, R. M. (1964). Estimates of survival from the sighting of marked animals. *Biometrika* 51, 429-438.

Coulson, J. C. (1960). A study of the mortality of the Starling based on ringing recoveries. *Journal of Animal Ecology* 29, 251-271.

Cox, D. R. (1972). *Analyse des Données Binaires* (Édition francaise). Paris: Dunod.

Fletcher, R. (1970). A new approach to variable metric algorithm. *Computer Journal* 13, 317-322.

Lebreton, J. D. (1977). Maximum likelihood estimations of survival rates from birds band returns: Some complements to age-dependent methods. *Biometrie-Praximetrie* 17, 145-161.

North, P. M. (1979). Relating Grey Heron survival rates to winter weather conditions. *Bird Study* 26, 23-28.

North, P. M. and Morgan, B. J. T. (1979). Modelling Heron survival using weather data. *Biometrics* 35, 667–681.

Sandland, R. L. and Kirkwood, G. P. (1981). Estimation of survival in marked populations with possibly dependent sighting probabilities. *Biometrika* 68, 531–541.

Seber, G. A. F. (1970). Estimating time-specific survival and reporting rates for adult birds from band returns. *Biometrika* 57, 313–318.

Seber, G. A. F. (1971). Estimating age-specific survival rates from bird-band returns when the reporting rate is constant. *Biometrika* 58, 431–497.

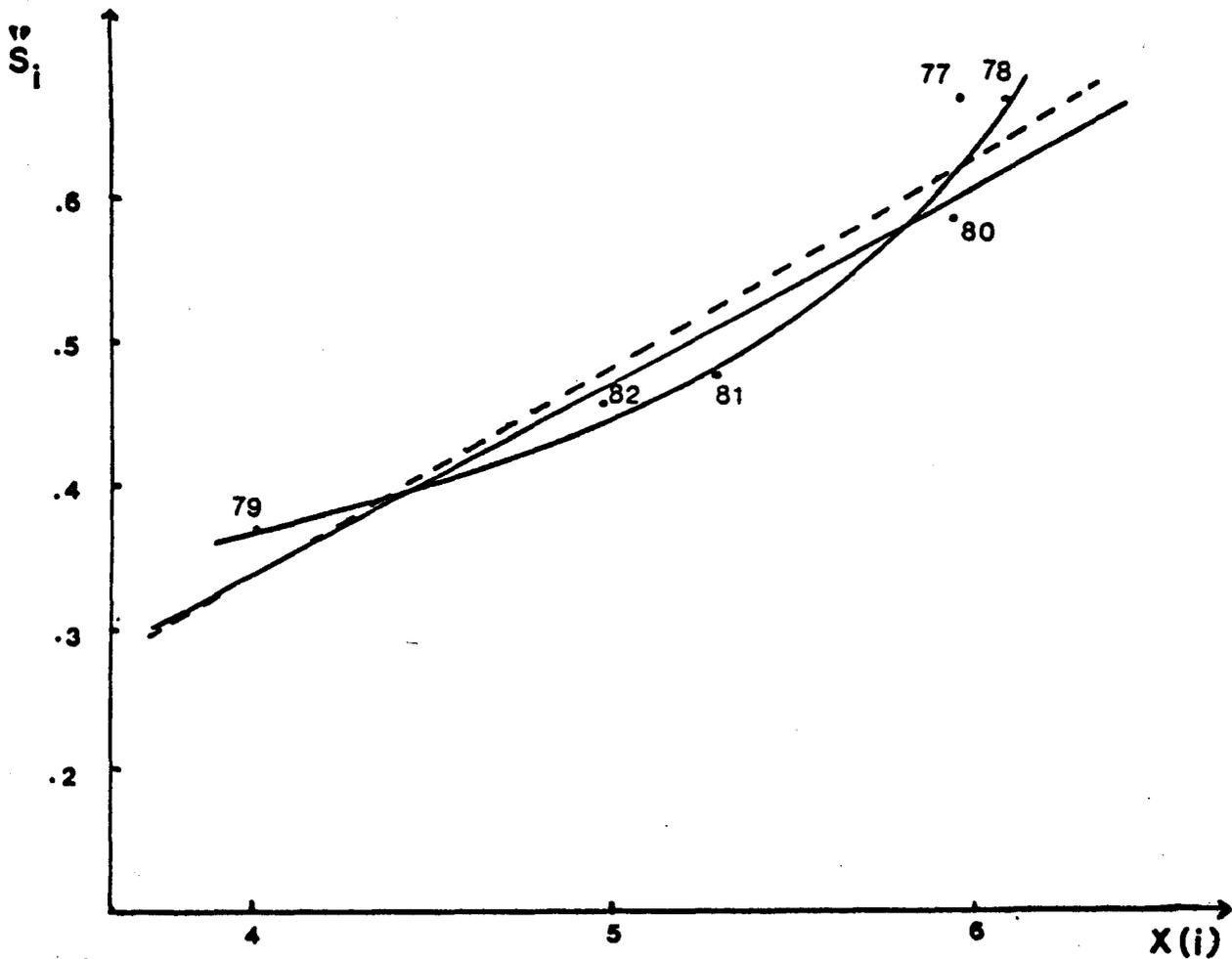Seber, G. A. F. (1973). *The Estimation of Animal Abundance and Related Parameters.* London: Griffin.

Figure 1. Relationship between the survival rate and the average winter temperature.

x(i) average winter temperature in °C from year i-1 to i.

Solid line: equation from model $\left(p, \; a_0 + a_1 x(t)\right)$ $\overset{..}{s}_i$ = -0.189(±0.119) + 0.133(±0.023)x(i)

Broken line: simple linear regression $\overset{..}{s}_i$ = -0.232(±0.123) + 0.143 (±0.015)x(i)

Solid curve: equation from model $\left(p, \; a_0 + a_1 x(i) + a_2 x^2(i)\right)$ $\overset{..}{s}_i$ =1.043 (±0.885)-0.365(±0.355)x(i)+0.049(±0.035)x^2(i)

(77 = winter 1976-1977; 78 = winter 1977-1978; etc.)

14